Title: Pupil Dilation During Encoding, But Not Type of Auditory Stimulation, Predicts Recognition Success in Face Memory

Authors: Sophie L. Cronin[1], Ottmar V. Lipp[2], Welber Marinovic[1]

Running Head: Pupil dilation and face encoding

[1] School of Population Health, Discipline of Psychology, Curtin University, Perth, Western Australia

[2] School of Psychology and Counselling, Queensland University of Technology, Brisbane, Queensland, Australia

Corresponding Authors: Welber Marinovic (welber.marinovic@curtin.edu.au)

.

Abstract

We encounter and process information from multiple sensory modalities in our daily lives, and research suggests that learning can be more efficient when contexts are multisensory. In this study, we were interested in whether face identity recognition memory might be improved in multisensory learning conditions, and to explore associated changes in pupil dilation during encoding and recognition. In two studies participants completed old/new face recognition tasks wherein visual face stimuli were presented in the context of sounds. Faces were learnt alongside no sound, low arousal sounds (Experiment 1), high arousal non-face relevant, or high arousal face relevant (Experiment 2) sounds. We predicted that the presence of sounds during encoding would improve later recognition accuracy, however, the results did not support this with no effect of sound condition on memory. Pupil dilation, however, was found to predict later successful recognition both at encoding and during recognition. While these results do not provide support to the notion that face learning is improved under multisensory conditions relative to unisensory conditions, they do suggest that pupillometry may be a useful tool to further explore face identity learning and recognition.

Keywords: pupillometry, faces, learning, encoding, arousal.

Introduction

Most of our experiences in daily life involve multiple sensory modalities. Predictably, our brains seem to have evolved to learn more efficiently from multisensory events (Kim et al., 2008; Seitz et al., 2006; Shams & Seitz, 2008). Here, we were interested in determining whether face recognition could be augmented by pairing faces with sounds. Face recognition is an important skill, and while familiar face recognition tends to be highly accurate, unfamiliar face recognition is error prone (e.g., Bruce et al., 1999; Clutterbuck & Johnston, 2005). Deficits in encoding novel faces, with limited opportunity to learn what features represent unique and identifying information result in poor later recognition and higher likelihood of misses. Understanding how to improve encoding conditions and identifying encoding conditions that result in better face recognition is an important part of understanding the mechanisms involved in this process.

Learning to recognise individual exemplars, in this case faces, requires the encoding of identity-specific information. Exploration of this topic with multisensory treatments suggests that multimodal learning can enhance learning relative to unimodal contexts (Shams & Sietz, 2008). Perceptual learning with audio-visual stimuli, for instance, can improve auditory learning. When learning to recognise voices, presenting faces and voices together can improve voice recognition relative to training with voices only (von Kriegstein & Giraud, 2006). However, congruence between modes in multimodal learning appears to be important with the relatedness of visual and auditory stimuli necessary for benefits over unimodal contexts (Kim et al., 2008). For instance, in a study where participants were shown a sequence of visual stimuli and had to respond each time an image was repeated, the presence or absence of a matching sound during the first presentation of the image affected their performance. Repetitions of stimuli (e.g., image of a bell) that were presented with congruent sounds (e.g., ‘dong’) in the first instance were better recognised than those presented with non-congruent sounds (e.g., ‘woof’) or no sound (Lehmann & Murray, 2005).

One way in which we can further explore the cognitive processes involved in face memory is to examine pupil dilation. Pupils dilate and constrict to control the amount of light entering the eye but these same changes in pupil size also occur in response to other physiological and cognitive changes. For instance, with increased cognitive effort, attention, and under conditions of heightened arousal, pupil dilation is altered (Kahneman & Beatty, 1966). It is generally believed that pupil size is correlated with memory strength and, consequently, the level of cognitive effort involved during encoding and confidence in later target recognition (Otero et al., 2011). For instance, Otero et al. (2011) found that when participants were given deep versus shallow encoding instructions, and reported stronger recognition of items, that pupil dilation was larger. Pupil size has also been found to be positively associated with the degree of specific detail recalled (Goldinger & Papesh, 2012). For example, Kucewicz et al. (2018) found that pupil size increased not only during encoding but also during recall, when no stimulus was presented, supporting the idea that pupil size is positively correlated with cognitive processes involved in memory encoding and retrieval.

This relationship between cognitive effort and pupil dilation is thought to reflect the connection of the pupil dilation system to the locus-coeruleus norepinephrine (LC-NE) network. The locus-coeruleus (LC) produces almost all of norepinephrine in the brain, a stress hormone and neurotransmitter contributing to various cognitive processes including attention and memory (Sara, 2009; Sara & Bouret, 2012). Neurons in the LC fire in two different modes, tonic and phasic, with tonic firing indicative of sustained baseline activity, and phasic firing indicative of transient responses to salient stimuli and cognitive processing (Aston-Jones et al., 2000). Task-evoked pupil responses can be explored to make inferences about phasic LC activation (Joshi et al., 2016; Kalwani et al., 2014).

Another factor that is sometimes important to memory and controlled by the LC-NE is arousal. Under conditions of high arousal, memory can be improved (e.g., Clewett et al., 2018). With increased arousal come signals of importance and relevance that can direct attention to enhance encoding. In the face recognition domain, this can be seen when emotional arousal is manipulated through emotional expressions, where out-group faces with emotional expressions are sometimes better remembered than those with neutral expressions (e.g., Cronin et al., 2019; Ackerman et al., 2006; Young & Hugenberg, 2012). Another example of this can be seen in memory formation, when particular features (i.e., scenes or objects) can be prioritised in memory when arousal is elicited through the threat of punishment (Clewett et al., 2018).

The effect of arousal, manipulated with auditory information, on learning and pupil dilation can be seen in other types of learning such as statistical learning. A study by Nassar et al. (2013) examined the activation of the LC-NE network and its effects on performance in a predictive-inference task using pupillometry as an index of LC activity. In the task, participants predicted an upcoming three-digit number. Subsequently, an outcome number was presented and participants then had to update their prediction for the next outcome, with new predictions constrained to be values between the current outcome and their previous prediction. Auditory cues were presented to indicate to participants to provide a new prediction but, on occasion, the sound was switched to a louder novel sound. The study found that pupil size changes associated with LC responses to novel sounds altered behaviour in the predictive-inference task. The direction of the effect was dependent on baseline pupil size: individuals with small baseline pupil size increased their learning rates significantly, whereas individuals with large pupil size at baseline showed a slight learning rate reduction. These results indicate that it is possible to alter cognitive processes using simple manipulations of sensory stimulation online, while participants engage in cognitive tasks. While in the Nassar et al. (2013) study sounds, which we meant to be ignored, were presented before participants had to make a prediction and it is not possible to guarantee that participants were not already engaging in mental calculations, we sought to replicate this online sensory stimulation manipulation of arousal in a face identity recognition study, presenting sounds before participants were asked to encode neutral faces.

The aim of the present study is to examine the effect of prior exposure to auditory stimulation on visual face memory and to explore associated changes in pupil dilation. In Experiment 1, an old/new face recognition task was conducted and during encoding, faces were presented in the context of no sound, low arousal sounds, or high arousal sounds. We expected that faces would be better remembered when encoding was preceded by high arousal sounds compared to no sound, or low arousal sounds. In Experiment 2, low arousal sounds were replaced with high arousal face relevant voice sounds. Here, we expected memory to be best when faces were encoded in the context of high arousal voice sounds, compared to non-voice high arousal sounds and no sound. Hits, confidence, and pupillometry were examined and explored.

Experiment 1

Method

Participants

Data from 38 undergraduate psychology students with normal or corrected-to-normal vision who participated for course credit were analysed (29 female, 9 male, M_{age}=21.08 years, SD_{age}=3.20 years, range=18-36 years). An additional 14 participants were excluded from analyses due to experimental completion errors (5) and chance or below chance recognition performance for encoded faces (9). Given the novel, exploratory nature of this study, sample size was guided by those of related studies exploring multimodal memory and pupil dilation (e.g., N=32 in Clewett et al., 2018; N=29 in von Kriegstein & Giraud, 2006; N=16, and N=11 in Lehmann & Murray, 2005) with oversampling. The study was conducted in accordance with the declaration of Helsinki. Ethics approval was obtained from Curtin University (HRE2018-0257). All participants provided written informed consent before starting the experiment.

Procedure

During the experiment, participants sat with their heads on a chin rest 57cm from a monitor screen. Participants completed an old/new recognition task comprising an encoding phase, filler task, and a recognition test. At the beginning of the encoding and recognition phases, the eye-tracker was calibrated using a nine-point calibration procedure. Participants were requested to rate valence and intensity of acoustic stimuli after the recognition phase.

Instruments and Apparatus

Visual stimuli were displayed on a 1920 x 1080 resolution computer monitor operating at 120 Hz and wore Bose headphones with active noise cancellation during tasks with sound and/or pupillometry. The experiment was programmed and run using MATLAB 2015b and PsychToolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). Pupillometry was conducted with an EyeLink 1000 Plus (SR Research, Ontario, Canada), sampling at a rate of 250Hz. Data from the right eye were analysed.

After the recognition phase, participants were presented with the sounds they had heard during the study and asked to rate them for perceived intensity[1] and valence. Intensity

[1] We do not consider arousal and intensity as synonymous, unidimensional constructs. For example, the processing of emotional prosody can be influenced by acoustic features expressing emotional arousal in a voice (Wiethoff et al., 2008). Although arousal and intensity have similar effects on pupil dilation, they are separate constructs and processed by

was rated on a scale anchored from 1 “not at all intense” to 7 “extremely intense. Valence was rated on a scale anchored from 1 “extremely unpleasant” to 7 “extremely pleasant”. Valence data did not produce any noteworthy results and so for brevity, only intensity ratings analyses will be presented.

Stimuli and Task

Ninety-six images of Caucasian young adult male and female faces with neutral expressions from the FACES database were used in the experiment (Ebner et al., 2010). Two images of each poser identity were used such that alternate images of encoded faces were presented at test. Images were 335 × 419 pixels in size. Luminance was equated using the MATLAB SHINE color toolbox (Dal Ben, 2021), achieving a pooled mean HSV of 0.46 (*SD*=0.24) across the images after correction. Images were converted to grayscale after luminance correction.

During encoding, participants viewed 24 faces one at a time (12 male, 12 female) on a grey screen, and were instructed to remember them as they would be asked to recognise them later. New, but highly similar photographs of the faces (see examples here: https://faces.mpdl.mpg.de/imeji/) were used during the test, but the participants were not informed of this. Trials progressed with an initial fixation point presented for 1500ms, followed by a face for 2000ms, and another fixation point for 1000ms. Interstimulus intervals were 4, 5, or 6 seconds long (random selection for each trial). The repeated measures factor for this experiment was sound type, with no sound, a low arousal sound, or a high arousal sound presented 500ms after the initial fixation point onset before face presentation. Eight faces were paired with each sound type, with face-sound pairings counterbalanced between-

different brain areas (see Anderson & Sobel, 2003), potentially accounting for unique variance in our data.

participants. The low arousal sound was a single 60dB dual-tone multi frequency (DTMF) signal with a sample rate of 8000 Hz that lasted 80ms (all faces in the low arousal sound condition were paired with the same tone). A Fast Fourier Transform of the DTMF signal was conducted to obtain its main frequency contents. The main frequencies identified were 943 Hz and 1333 Hz. There were eight unique sounds in the high arousal condition which consisted of complex sounds (e.g., animal, object, and arcade sounds)[2] sourced from online databases. These sounds were presented once each and had a peak of approximately 80dB and duration varied between 429-1312ms. Pupil dilation was measured on each trial from initial fixation onset to post-fixation offset. We selected sounds for our low and high arousal conditions based on their known effects on pupil dilation. As reviewed by Zekveld et al. (2018), novelty and loudness can both independently influence pupillary responses, with more varied and loud sounds leading to larger changes in dilation. In contrast, repeated sounds tend to lead to reduced dilation over time (Zekveld et al., 2018). We chose our sounds with these factors in mind in order to elicit the largest possible difference in pupil dilation between conditions.

After encoding, participants completed a filler task for approximately five minutes where they completed a series of simple word puzzles. Following this, participants completed a recognition test. The 24 faces viewed in encoding, and an additional 24 unseen faces (12 male, 12 female) were presented one at a time. New images were presented for all faces (alternate images of the poser identity were used for encoded faces, with learnt and tested images counterbalanced between-participants). Faces that were seen during encoding and unseen during recognition were counterbalanced between-participants. An initial fixation

[2] Sounds were sourced from splice.com ('SuperheroGadgetOn_HV.823', 'SteamWhistle_BW.60898', 'HRD_Pause_Game', 'FF_ES_foley_rattlesnake_brown', 'ESM_Explainer_Video_One_Shot_Foley_Bike_Bicycle_Bell_Ding_1_Dry', 'BirdBlueJayScreech_BWU.657', '018_FX_-_Zenhiser_TD') and Pixabay.com ('Mario coin 200bpm').

point was presented for 2000ms, followed by a face for 2000ms, a post-fixation point for 1500ms, and then two question prompts. Participants were asked if they had seen the face earlier during the experiment (yes/no), and how confident they were in their answer (responses were given on a 0-100 scale from “not at all confident” to “completely confident”), responding using the keyboard. Pupil dilation was measured on each trial from fixation onset to post-fixation offset.

Data Processing and Analysis

Only test scores from encoded faces were analysed[3]. Test response data are reported as hits (0 = did not remember the face, 1 = remembered the face). Pupil metrics were extracted from the encoding and recognition phases of the experiment. Initial processing of the data removed blinks identified by the EyeLink 1000 Plus and replaced missing data with linear interpolation using the “rollapply” function (five-point moving average) from the *Zoo* package in R. In cases where missing and/or interpolated data accounted for 20% or more of the trial, the entire trial was excluded (Nyström et al., 2013). Where visual inspection of the data identified additional measurement artifacts (e.g., partial blinks), the trial was also excluded. Pupil dilation in all trials was baseline (absolute pupil size) corrected using the subtractive method (dilation = absolute pupil size for each point of the time series – the mean absolute pupil size during baseline), and we inspected absolute baseline values to make sure that no unusually small (<500 arbitrary units) pupil sizes affected our results as suggested by Mathot et al. (2018). From each phase of the experiment, average dilation during baseline (500ms following initial participant fixation – absolute values), average dilation during face

[3] A GLMM with accuracy as a fixed factor and participant ID and face as random factors found there was no differences in accuracy between encoded faces and new faces in either Experiment 1 ($\chi^2(1)=1.63$, $p=.202$) or Experiment 2 ($\chi^2(1)=0.01$, $p=.918$).

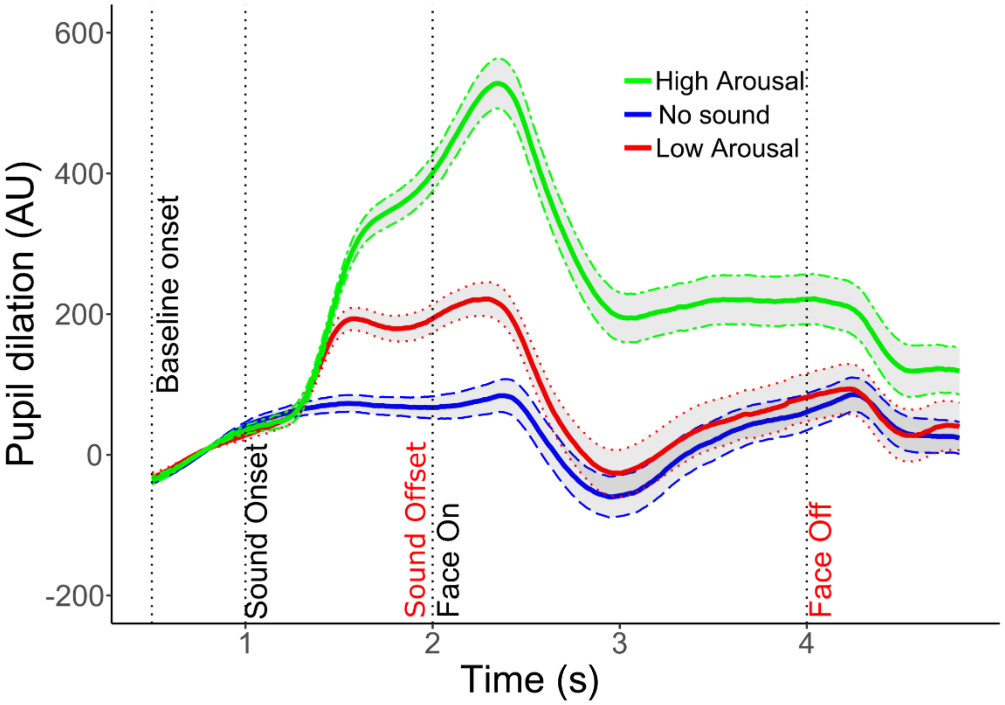
presentation, and time to peak dilation during face presentation were recorded. These pupil metrics, along with confidence scores, were standardised by participant ID for analysis.

Logistic generalised linear mixed model (GLMM) and linear mixed model (LMM) analyses were conducted on the hits and confidence data, respectively. These models included participant ID and face as random factors and sound type as the fixed factor. As reported in the results section, exploratory analysis included eye metric measurements obtained both during encoding and recognition as continuous covariates, these included baseline pupil size, time to peak dilation during face presentation, and average pupil dilation during face presentation. We also used LMM to examine changes in pupil dilation during face presentation at encoding as a function of sound type (fixed factor). Satterwaithe's (1946) method was used to estimate degrees of freedom. Analyses not including pupil metrics used all cases (38 participants) and observations (912 observations) except for analyses that included intensity ratings for which there was some missing data due to experimental error (resulting in 848 observations from 38 participants). Analyses including pupil metrics (after exclusions were applied) consisted of 681 of 912 observations from 37 of 38 participants.

Results

The LMM analysis examining the effect of sound type (no sound, low arousal, high arousal) on average pupil dilation during face presentation at encoding confirmed the manipulation worked as intended. As shown in Figure 1, average pupil dilation during face presentation at encoding was larger after the low (β=53.71, SE=26.36, 95%CI[2.05, 106.38], *t*(648)=2.04 *p*=.04) and high (β=268.75, SE=26.94, 95%CI[215.95, 321.54], *t*(649)=9.98 *p*<.001) arousal sounds, compared to after no sound (*F*(2,648)=55.15, *p*<.001).

Figure 1: Averaged pupil response and standard error across participants as a function of sound type in Experiment 1.



A GLMM analysis was conducted on the hit data with sound type as a fixed factor to determine if sound type predicted hits (see Table 1). Results indicated that the type of sound presented during encoding did not influence whether the face was correctly recognised (χ^2(2)=0.0002, *p*=.99). A LMM examining the effect of sound type on confidence data (see Table 1) similarly, found that sound type also did not predict confidence at test (*F*(2,769)=1.71, *p*=.18). Exploratory GLMM and LMM analyses using perceived intensity and valence ratings of each sound as fixed factors (see Table 1), instead of their sound type category, found no effect of the manipulation on hits (Intensity: χ^2(1)=0.14, *p*=.70; Valence: χ^2(1)=0.41, *p*=.52) or confidence (Intensity: *F*(1,816)=2.11, *p*=.14; Valence: *F*(1,794)=0.08, *p*=.77).

Table 1

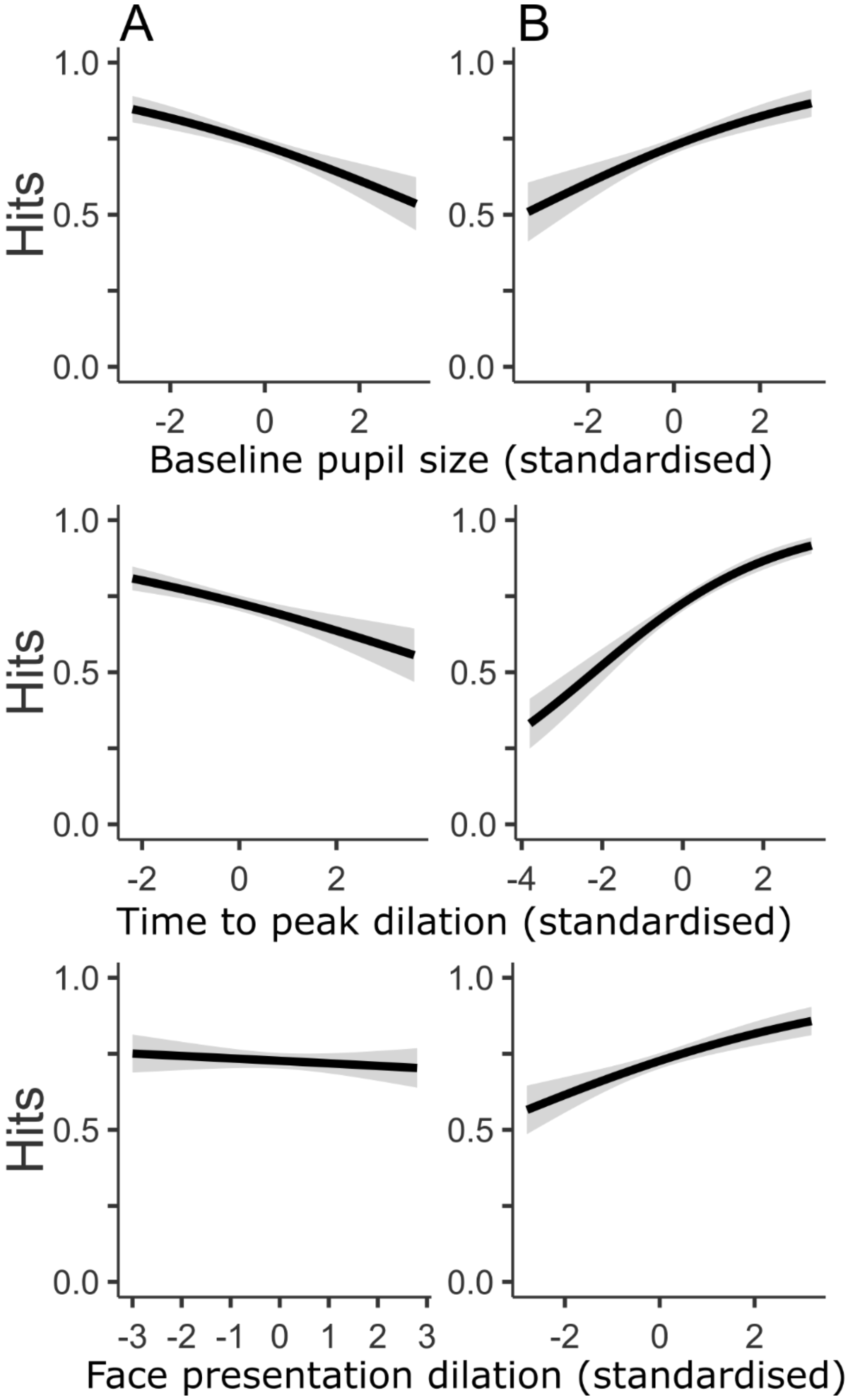
Hit rate, mean confidence, mean perceived intensity, and mean valence for each sound type in Experiment 1

Score	No Sound	Low Arousal	High Arousal
Hits	0.72(0.03)	0.72(0.03)	0.72(0.03)
Confidence	63.08(2.03)	66.2(2.04)	66.3(2.03)
Intensity		3.18(0.11)	4.01(0.12)
Valence		3.68(0.09)	4.16 (0.08)

Note. Values are estimated marginal means of the models. Standard errors are reported in brackets.

While sound type did not predict hits at test, an exploratory GLMM analysis on the hits data with pupil metrics included as fixed factors, found that pupil dilation at encoding and recognition did. Smaller baseline pupil size ($\chi^2(1)=6.08$, $p=.01$, β =-0.26, SE=0.11, 95%CI[0.63, 0.95]), and earlier peaks in dilation during face presentation ($\chi^2(1)=4.88$, $p=.02$, β =-0.21, SE=0.09, 95%CI[0.67, 0.98]) at encoding predicted a higher likelihood of correctly recognising the presented face later. Average pupil dilation during face presentation at encoding was not related to hits ($\chi^2(1)=0.163$, $p=.68$). Pupil dilation at recognition also predicted performance, however, conversely to encoding, larger baseline pupil size ($\chi^2(1)=6.31$, $p=.01$, β=0.28, SE=0.11, 95%CI[1.06, 1.64]), later peaks in dilation during face presentation ($\chi^2(1)=21.30$, $p<.001$, β=0.44, SE=0.10, 95%CI[1.29, 1.88]), and larger average dilation during face presentation ($\chi^2(1)=5.42$, $p=.020$, β=0.26, SE=0.11, 95%CI[1.04, 1.60]) predicted a higher likelihood of a hit.

Figure 2: *Likelihood of a hit as predicted by encoding (A) and recognition (B) pupil metrics in Experiment 1.*



Note. Standard error plotted in grey.

In addition to predicting hits, pupil dilation during recognition predicted confidence in responses. A LMM conducted on confidence data with pupil metrics included as fixed factors found that larger baseline pupil dilation (*F*(1,643)=13.03, *p*<.001, β=3.77, SE=1.05, 95%CI[1.72, 5.82]), later peaks during face presentations (*F*(1,640)=10.51, *p*<.001, β=3.41, SE=1.05, 95%CI[1.35, 5.48]), and larger average dilation during face presentation (*F*(1,626)=26.83, *p*<.001, β=4.81, SE=0.93, 95%CI[2.99, 6.63]) were predictive of higher confidence in responses. Encoding pupil dilation metrics were not related to confidence (all *F* <.30, all *p*>.59).

Discussion

The results of Experiment 1 did not support our hypothesis that auditory stimulation would improve face recognition. Sound type had no effect on hits, or confidence in response at test. One potential reason for this may be that sounds were unrelated to, or incongruent with, the face stimuli. A recent meta-analysis conducted by Li and Deng (2022) found that multisensory stimulation can either facilitate or hinder learning depending on the congruency of the stimuli; congruent stimuli tend to facilitate learning, while incongruent stimuli can interfere with it (see also Lehmann & Murray, 2005). Therefore, it is plausible that beyond just increasing physiological arousal before viewing the face, some binding or congruency (e.g., faces and voices) between the visual and auditory information may be necessary to improve encoding. To explore this possibility, Experiment 2 replicated Experiment 1 with the inclusion of voice sounds. As there was no difference between the low arousal and high arousal conditions on performance, Experiment 2 did not include low arousal sounds. If sound relevance is needed for sound cues to improve memory, then we expected that voice sounds would improve performance compared to the high arousal (object) sounds and no sound.

Exploration of pupillometry data produced some interesting results that were also followed up in Experiment 2. Results suggest that smaller baseline pupil dilation and earlier peak in pupil dilation while the face is presented at encoding increase the likelihood of a hit. This unexpected finding was examined again in Experiment 2 to see if this association replicates.

Experiment 2

Method

Participants

Data from 36 undergraduate psychology students with normal or corrected-to-normal vision who participated for course credit were analysed (22 female, 14 male, M_{age}=21.11 years, SD_{age}=3.69 years, range=18-39 years). An additional 11 participants were excluded from analyses due to experimental completion errors (1) or chance and below chance recognition performance for encoded faces (10).

Procedures

All procedures were the same as Experiment 1.

Instruments and Apparatus

The instruments and apparatus used were identical as for Experiment 1.

Stimuli and Task

All details were the same as Experiment 1 exempting the following changes to stimuli. The low arousal sound condition was replaced with a voice condition in which eight high arousal, unique speech sounds were used. Each sound consisted of a different person speaking a different statement (e.g., “It’s eleven o’clock”) and were adjusted to peak at approximately

80dB. Stimuli were sourced from the CREMA-D (Cao et al., 2014) and RAVDESS (Livingstone & Russo, 2018) databases[4]. Sounds were randomly paired with faces during encoding; however, sex of the speaker and sex of the face were always matched. All sound stimuli were standardised to 1000ms in duration by adjusting tempo, and/or repeating the sound (four instances in the high arousal sound condition required a repetition). This resulted in the offset of all sounds coinciding with the onset of the face images.

Data Analysis and Processing

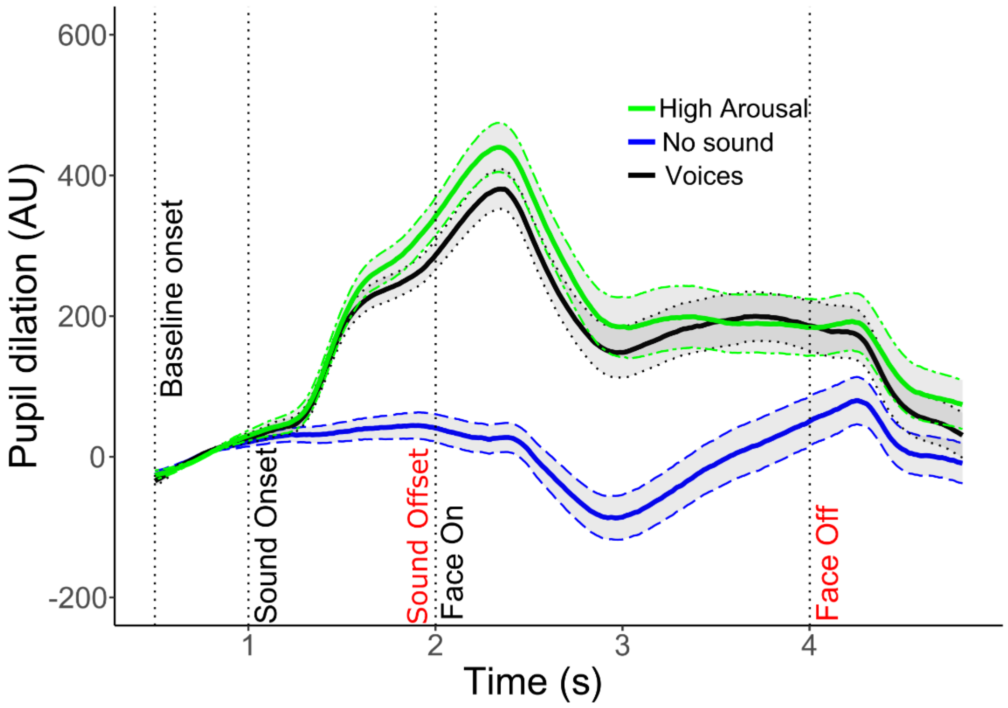
Data were analysed and processed as in Experiment 1. Analyses not including pupil metrics used data from all 36 participants, and 864 observations (those including ratings consisted of data from 36 participants and 848 observations). Analyses including pupil metrics consisted of data from 36 participants, and 610 observations. GLMM analyses were conducted on the hit data, LMM analyses on the confidence data, with participant ID and face included as random factors in all analyses.

Results

An LMM analysis examining the effect of sound type (no sound, high arousal, voice) on average pupil dilation during face presentation at encoding confirmed the manipulation worked as intended. As depicted in Figure 3, average pupil dilation during face presentation at encoding was larger after the high arousal (β=265.68, SE=27.08, 95%CI[212.60, 318.75], *t*(553)=9.81, *p*<.001) and voice (β=241.89, SE=27.72, 95%CI[187.56, 296.21], *t*(550)=8.73, *p*<.001) sounds, compared to after no sound (*F*(2,555)=57.92, *p*<.001).

[4] Items used were: '1002_DFA_NEU_XX', '1003_TAI_NEU_XX', '1009_IWL_NEU_XX', '1010_TIE_NEU_XX', '1011_TSI_NEU_XX', '1014_IOM_NEU_XX', '1023_MTI_NEU_XX', '1033_IEO_NEU_XX'.

Figure 3: Averaged pupil response and standard error across participants as a function of sound type in Experiment 2.



A GLMM analysis of the hits data with a fixed factor of sound type was conducted (see Table 3). Results indicated that, as in Experiment 1, the type of sound presented during encoding did not influence whether the face was correctly recognised at test ($\chi^2(2)=1.30$, $p=.52$). A LMM on confidence data and a fixed factor of sound type (see Table 3) similarly, found that sound type also did not predict confidence at test ($F(2,664)=0.04$, $p=.95$). GLMM analyses with perceived intensity and valence ratings as factors (see Table 3) of each sound found that higher perceived intensity predicted higher likelihood of hits ($\chi^2(2)=4.09$, $p=.04$, β=0.07, SE=0.04, 95%CI[1.00, 1.16]), but valence had no effect ($\chi^2(2)=4.09$, $p=.42$, β=0.03, SE=0.03, 95%CI[0.95, 1.11]). Similar analyses of confidence data found that neither perceived intensity nor valence predicted confidence (Intensity: $F(1,812)=0.04$, $p=.83$; Valence: $F(1,801)=1.27$, $p=.25$).

Table 2

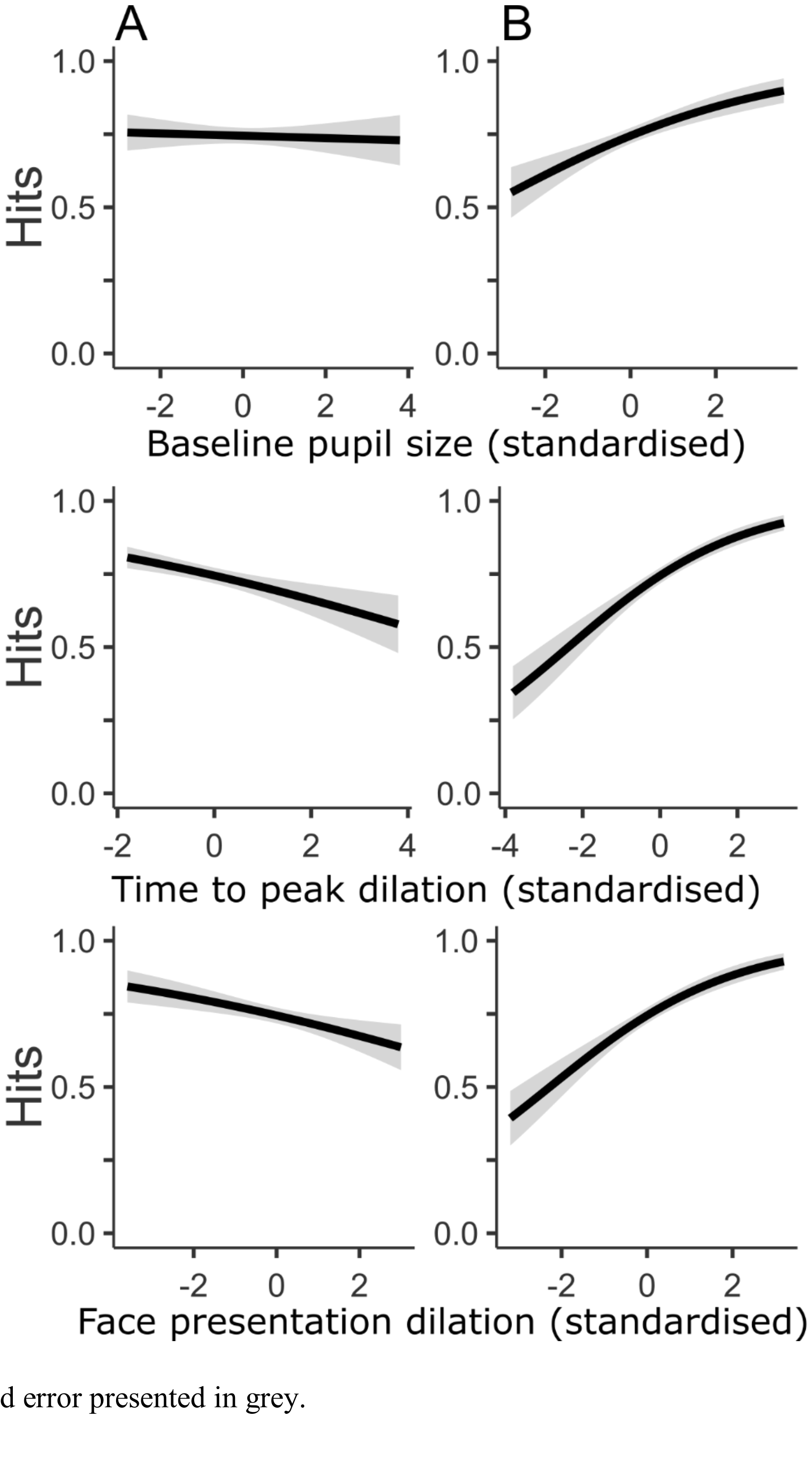
Hit rate, mean confidence, mean perceived intensity, and mean valence for each sound type in Experiment 2

Score	No Sound	High Arousal	Voice
Hits	0.69(0.03)	0.73(0.03)	0.73(0.03)
Confidence	61.47(2.44)	61.41(2.42)	62.06(2.45)
Intensity		4.35(0.14)	2.80(0.14)
Valence		3.32(0.09)	4.09(0.09)

Note. Values are estimated marginal means of the models. Standard errors are reported in brackets.

A GLMM analysis on hit data with pupil metrics included as fixed factors found that pupil dilation at encoding and recognition predicted hits. An earlier peak in dilation during face presentation (χ^2(1)=3.84, *p*=.05, β=-0.20, SE=0.10, 95%CI[0.67, 1.00]) at encoding predicted a higher likelihood of correctly remembering the presented face later. However, unlike Experiment 1 baseline pupil size at encoding did not predict hits (χ^2(1)=0.04, *p*=.85). Average dilation during face presentation (χ^2(1)=2.63, *p*=.10) at encoding was also not related to hits. Pupil size at recognition predicted performance in the same fashion as in Experiment 1, with larger baseline pupil size (χ^2(1)=6.83, *p*=.009, β=0.31, SE=0.12, 95%CI[1.08, 1.72]), later peaks in dilation during face presentation (χ2(1)=18.79, *p*<.001, β=0.45, SE=0.10, 95%CI[1.28, 1.93]), and larger average dilation during face presentation (χ^2(1)=15.31, *p*<.001, β=0.47, SE=0.12, 95%CI[1.26, 2.03]) predicting a higher likelihood of a hit.

Figure 4: *Likelihood of a hit as predicted by encoding (A) and recognition (B) pupil metrics in Experiment 2.*



Note. Standard error presented in grey.

A LMM of confidence data with pupil metrics as fixed factors showed that pupil dilation during recognition also predicted confidence, replicating the pattern of Experiment 1. Larger baseline pupil size (*F*(1,528)=8.78, *p*<.01, β=3.68, SE=1.24, 95%CI[1.25, 6.11]), later peaks during face presentations (*F*(1, 570)=12.47, *p*<.001, β=3.95, SE=1.12, 95%CI[1.76, 6.15]), and larger average dilation during face presentation (*F*(1,556)=30.35, *p*<.001, β=6.87, SE=1.25, 95%CI[4.42, 9.31]) were predictive of higher confidence in responses. Encoding pupil dilation metrics were not related to confidence (all *F* <1.93, all *p*>.16).

Discussion

Replicating Experiment 1, sound type did not influence recognition memory performance. However, unlike in Experiment 1, the perceived intensity of sound stimuli, irrespective of sound type, did predict hits. The more intense the sound was perceived to be, the more likely it was for that face to be recognised at test. One possible explanation for this is that Experiment 2 included more 80dB sound trials and may have increased power to observe this relationship. Examining the results from analyses using pupil metrics we can see that all recognition results replicated from Experiment 1. Unlike Experiment 1, baseline pupil size at encoding was no longer related to likelihood of hits, though as before, earlier peaks in pupil dilation during face presentation were related to an increase in likelihood of a hit. In the next section we combine the datasets from Experiment 1 and 2 to examine whether a larger dataset could provide additional insights about face encoding due to increased statistical power.

Combined Analyses of Experiments 1 and 2

A GLMM analysis conducted on the overall hits data with sound type and experiment as fixed factors indicated that the type of sound presented during encoding did not influence whether the face was correctly recognised at test (χ^2(3)=0.71, *p*=.87), nor were there differences across the experiments (χ^2(1)=0.05, *p*=.81). With regards to the effect of sound type on face learning, our primary interest in this study, a follow-up Bayesian analysis using the *generalTestBF* function from the R package “BayesFactor” found decisive evidence for the null hypothesis (B_{01} = 626). A LMM examining confidence data with sound type and experiment as fixed factors similarly, found that sound type also did not predict confidence at test (*F*(3,1466)=0.59, *p*=.62), nor did experiment (*F*(1,91)=1.22, *p*=.19). LMM analysis of the effect of perceived intensity ratings found these did not significantly predict hits (χ^2(1)=1.20, *p*=.27), or confidence (*F*(1,1650)=0.37, *p*=.54) in the combined dataset.

GLMM analysis of the effect of pupil metrics found that pupil dilation at encoding and recognition predicted hits. A smaller baseline pupil size (χ^2(1)=5.16, *p*=.02, β=-0.17, SE=0.08, 95%CI[0.72, 0.98]) and earlier peak in dilation during face presentation (χ^2(1)=8.55, *p*<.01, β=-0.20, SE=0.07, 95%CI[0.71, 0.94]) at encoding predicted a higher likelihood of correctly remembering the presented face later. Average dilation during face presentation (χ^2(1)=1.86, *p*=.17, β=-0.10, SE=0.07, 95%CI[0.78, 1.04]) at encoding was not related to hits. Pupil dilation at recognition also predicted performance, however, conversely to encoding, larger baseline pupil size (χ^2(1)=11.75, *p*<.001, β=0.27, SE=0.08, 95%CI[1.12, 1.54]), later peaks in dilation during face presentation (χ^2(1)=40.46, *p*<.001, β=0.44, SE=0.07, 95%CI[1.36, 1.78]), and larger average dilation during face presentation (χ^2(1)=16.93, *p*<.001, β=0.33, SE=0.08, 95%CI[1.19, 1.63]) predicted a higher likelihood of a hit.

In addition to predicting hits, pupil dilation during recognition predicted confidence. Larger baseline pupil size (*F*(1,1220)=21.77, *p*<.001, β=3.74, SE=0.80, 95%CI[2.17, 5.31]), later peaks during face presentations (*F*(1,1209)=36.57, *p*<.001, β=4.33, SE=0.72, 95%CI[2.93, 5.73]), and larger average dilation during face presentation (*F*(1,1217)=36.59, *p*<.001, β=4.88, SE=0.81, 95%CI[3.30, 6.46]) were predictive of higher confidence in responses. Encoding pupil dilation metrics were not related to confidence (all *F* <1.55, all *p*>.30).

General Discussion

The primary aim of this study was to determine if auditory cues prior to encoding would improve face memory. The results of Experiments 1 and 2 suggest that it does not. In both experiments, the type of sound presented during encoding had no effect on whether the face would be later remembered. In fact, a follow-up Bayesian analysis provided decisive support for the null hypothesis. There was some suggestion that the intensity of the sound mattered in Experiment 2, with increasing perceived intensity of the sound predictive of a higher likelihood of a hit, however, this relationship was not present in Experiment 1 and did not hold in the combined analysis. There was also no effect of sound type on confidence, suggesting it is not the case that memory strength is being influenced by sound but just not translating to different performances. It seems then that presenting sounds immediately prior to encoding, whether low arousal, high arousal, or visual stimulus congruent, does not improve face recognition under the scenario tested in our experiments.

While our studies did not show any performance improvements associated with the sounds presented, we are confident that our manipulation was successful in engaging LC activity. Our claim is based on evidence from studies that have recorded from neurons in rhesus monkeys to examine the relationship between pupil size and neural activity in LC. Recording from neurons across LC, superior and inferior colliculus, and cingulate cortex in rhesus monkeys, Joshi et al. (2016) found that pupil size was associated with spiking activity across all regions. Their results showed that spiking activity in LC was not only predictive of pupil size changes but also preceded activity in all other brain areas. In addition, a recent study by Megemont et al. (2022) showed the same positive and monotonic relationship between LC spiking activity and pupil size. Although these later results showed substantial variability for small pupil changes, they demonstrated that large pupil dilation changes – >2SD and similar to the changes we induced with our sounds – were accurately predictive of

LC spiking in real-time. Therefore, we believe it is safe to assume that our sound manipulation was effective in modulating the release of NE throughout the cortex. In this case, LC-NE could have affected cognitive processes via alternations in the gain of neurons and/or cognitive shifts induced by arousal changes (Sara, 2009; Sara & Bouret, 2012).

The role of LC activity in modulating cognitive function, in particular its relationship to arousal-related memory processing, relates to adaptive gain control theory. Adaptive gain control theory proposes that the brain adjusts the gain of its neural responses in order to optimize the processing of sensory information (Aston-Jones & Cohen, 2005). For example, when an animal is confronted with a loud noise, its brain may increase the gain of its auditory responses in order to better process the information contained in the sound, enhancing the animal's ability to remember the event. In a similar vein, quick changes in arousal induced by phasic LC activity can reorganise the functional connectivity of neural networks and facilitate some behaviours (Sara, 2009; Sara & Bouret, 2012). Irrespective of the specific mechanisms by which cognitive function was expected to be modulated by sounds in our task, we found no evidence this occurred. In what follows, we consider why this was the case.

In a study by Jacobs et al. (2020), the authors used 7T fMRI to examine the dynamic changes in LC during arousal-related memory processing in humans. Although they found that LC activation plays a role in controlling arousal, LC activity alone was not sufficient to form memories of emotional experiences: these experiences required interactions with the amygdala for effective memory formation. Similar to our study, Jacobs et al. (2020) used face stimuli in a face-name association task. However, their stimuli had faces expressing different emotions (high arousing and low arousing) and, therefore, the faces themselves evoked changes in arousal. It is possible that although we were successful in activating LC, our manipulation was not successful to engage the required interaction with the amygdala.

The design of our study was inspired by the study of Nassar et al. (2013) which found that novel sounds, which provided no information about the upcoming stimuli, induced pupil size changes that predictably altered the influence of new data. In other words, statistical learning was modulated by the effect of task-unrelated novel sounds on LC activity. Similarly, in Experiment 1, we sought to induce pupil size changes just before a face was presented by presenting task-unrelated sounds, but unlike Nassar et al. (2013) our manipulation failed to produce the expected results. One potential explanation for the failure of Experiment 1 is that our task may require a clear congruency between visual and auditory stimuli (e.g., face and voice *vs.* face and a bell ringing). To tackle this issue, in Experiment 2 we paired faces and voices but again found no benefit of high arousal auditory stimuli on face learning. Note, however, that the auditory and visual stimuli were still offset, with voice sounds concluding at the onset of faces. It may be that the binding we hoped to achieve was not successful as there was no physical overlap between sound and visual stimulus, though sounds clearly increased arousal during face learning. A study by von Kriegstein and Giraud (2006) found that voice recognition was improved in the audio-visual context, when voices to be learnt were followed by a face. However, their design had participants view voice and face stimuli both sequentially and simultaneously as audio-visual stimuli. As such, employing simultaneous presentation of auditory and visual stimuli in our design may have more success in demonstrating the influence of sound on face memory.

Another possibility is that memory encoding is not influenced by sound presentation in the same way that statistical learning is. It might be the case that feedback presented in the Nassar et al. task on a trial-by-trial basis is important to the mechanism and that enhancing arousal with sounds allows participants to better learn from feedback (see also Leow et al., 2021). In the case of an old/new recognition task, there is no feedback, and so participants cannot change their strategy on a trial-by-trial basis. Likewise, the effect of this manipulation may be enacted on decision making processes at test, rather than on encoding processes. Another difference between studies noted by one of our reviewers is that Nassar et al. (2013) observed the effect of sounds immediately after their presentation, whereas in our study the effects of sounds on face learning were only tested later, during the recognition phase of our study. So, it is plausible that these effects are short lived and difficult to measure after a relatively long interval (> 5 min.). Alternatively, it is possible that the effects of sound on memory formation might take even longer to manifest. In a recent study pairing motor actions and loud acoustic stimuli, Leow and colleagues (2021) found that accessory sounds did not affect performance during acquisition but effects on retention were observed after an overnight delay (>17 hours). Therefore, it remains to be tested whether sounds could improve face learning if a longer consolidation interval (overnight) was allowed (see McGaugh, 2000, 2018).

While we did not find sound type to affect memory, exploratory analyses of the pupil dilation data yielded some promising results. During encoding, pupil dilation predicted later hits. When viewing the face, peak pupil dilation latency was negatively associated with performance in Experiments 1 and 2. In Experiment 1, and the overall analysis, average baseline pupil size before any stimuli were presented on a trial was also predictive of later performance, with smaller pupil diameter related to higher likelihood of a hit for that face. We did not, however, find evidence that any encoding pupil dilation metrics were related to confidence in responses.

These encoding pupil dilation results add to the mixed findings in the literature. While some studies have found results like ours that there is a negative relationship between encoding pupil dilation and recognition memory performance (Wetzel et al., 2020; Naber et al., 2013; Kafkas & Montaldi, 2011; Kafkas, 2021), others have found the opposite, observing a positive relationship (Miller & Unsworth, 2019; Papesh et al., 2012; Kafkas &

Montaldi, 2015). Many of these studies, however, were interested in effects on memory type and so when exploring these relationships, made distinctions between faces that are forgotten, remembered with a sense of familiarity only, or remembered by recollection. Our data do not include this distinction of memory type, however, encoding pupil dilation did not relate to confidence in our analyses suggesting that memory strength was not related to it. It is important to note that unidimensional and dual process models of memory disagree on whether familiarity and recollection represent merely different strengths of memory (low and high, respectively) or distinct memory processes and so comparisons between our result and the results of these previous studies should be done with caution.

Studies that have found negative relationships between encoding pupil dilation and memory have varied in assessed pupil metrics and study design. Unlike our intentional encoding design, they have used incidental encoding procedures where participants were surprised with a memory test. Additionally, studies found negative relationships between memory and pupil dilation using different metrics, focusing on average pupil dilation during stimulus presentation and examining pupil dilation and constriction behaviour. This is distinct from our findings that average dilation is not related to performance, but instead peak latency is.

That baseline pupil size during encoding was negatively related to performance is a novel finding. As previously mentioned, Nassar et al. (2013) found that baseline pupil size was related to learning rates, however, in an ‘inverted U’ pattern, indicating best performance at intermediate pupil dilations. While our manipulation of sound type was not observed to alter memory performance, the expectation of loud sounds throughout the encoding session may have changed participants’ readiness to learn faces. It may be that those who had lower baseline pupil size were better equipped to tolerate conditions of enhanced arousal and not become overstimulated whilst encoding the faces. Similarly, those whose peak in dilation during face presentation was earlier may have been better able to recover from arousal increases faster, and better focus on encoding faces successfully. From a physiological perspective, these baseline effects might be related to the tonic firing of LC neurons, which is believed to reflect global brain states such as arousal (Aston-Jones & Cohen, 2005). Therefore, although we failed to modulate face learning via the phasic activation of LC, our results indicate that it can be informative to monitor the dynamic changes in pupil size as a proxy for changes in brain states over time.

Another explanation from recent evidence is that unexpected and expected novelty can drive encoding dilation – memory relationships. Kafkas (2021) examined the relationship between pupil dilation across encoding and memory success and memory type, and found that when novelty was unexpected, there was a positive relationship between encoding pupil dilation and later memory. On the other hand, expected novelty produced a negative relationship between encoding pupil dilation and memory. As participants in our study were informed of the presence of sounds during encoding, they may have been prepared for the stimuli leading to reduced novelty responses that may otherwise further enhance pupil dilation. There are, of course, other ways to explain the negative relationship between encoding pupil dilation latency and hits that we cannot exclude. For instance, smaller pupil dilation might be indicative of the absence of cognitive processes that reduce memory performance such as off-task thinking (Unsworth & Robison 2017).

Pupil dilation during recognition was also predictive of performance and confidence. Increased baseline pupil dilation, peak pupil dilation latency during face viewing, and average dilation during face presentation, were all associated with a higher likelihood of a hit on that trial. This is consistent with other literature that suggests that processes of recognition or retrieval of memories involve more cognitive processing and that this can be observed in pupil dilation (Brocher & Graf, 2016; Otero et al., 2011; Papesh et al., 2012; Vo et al., 2008;

Elphick et al., 2020). Likewise, stronger memories in which observers have more confidence have been associated with larger pupil dilation in auditory recognition (Papesh et al., 2012).

In conclusion, we found that presenting sounds prior to the presentation of faces did not improve their later recognition. Nonetheless, pupillometry provided several pieces of information regarding the dynamics of face learning. Pupillometry is a useful tool in exploring the encoding and recognition of face identity, providing vital information about the dynamic changes in brain states.

Funding: This work was supported by the Australian Research Council [DP180100394].

References

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., . . . Schaller, M. (2006). They all look the same to me (unless they're angry): From out-group homogeneity to out-group heterogeneity. *Psychological Science, 17*(10), 836-840. https://doi.org/10.1111/j.1467-9280.2006.01790.x

Anderson, A. K., & Sobel, N. (2003). Dissociating Intensity from Valence as Sensory Inputs to Emotion. *Neuron*, *39*(4), 581-583. https://doi.org/https://doi.org/10.1016/S0896-6273(03)00504-X

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

Aston-Jones, G., Rajkowski, J., & Cohen, J. (2000). Locus coeruleus and regulation of behavioral flexibility and attention. *Progress in Brain Research, 126*, 165-182. https://doi.org/10.1016/S0079-6123(00)26013-5

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436

Brocher, A., & Graf, T. (2016). Pupil old/new effects reflect stimulus encoding and decoding in short term memory. *Psychophysiology, 53*(12), 1823–1835. https://doi.org/10.1111/psyp.12770

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339.

Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset. *IEEE* *transactions on affective computing, 5*(4), 377–390. https://doi.org/10.1109/TAFFC.2014.2336244

Clewett, D. V., Huang, R., Velasco, R., Lee, T., & Mather, M. (2018). Locus coeruleus activity strengthens prioritized memories under arousal. *Journal of Neuroscience, 38*(6), 1558-1574. https://doi.org/10.1523/JNEUROSCI.2097-17.2017

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17*(1), 97-116.

Cronin, S. L., Craig, B. M., & Lipp, O. V. (2019). Emotional expressions reduce the own-age bias. *Emotion, 19*(7), 1206-1213. http://doi.org/10.1037/emo0000517

Dal Ben, R. (2021). SHINE_color and Lum_fun: A set of tools to control luminance of colorful images (Version 0.3). [Computer program]. https://doi.org/10.17605/OSF.IO/AUZJY

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351-362. https://doi.org/10.3758/BRM.42.1.351

Elphick, C. E. J., Pike, G. E., & Hole, G. J. (2020). You can believe your eyes: Measuring implicit recognition in a lineup with pupillometry. *Psychology, Crime & Law, 26*(1), 67-92. https://doi.org/10.1080/1068316X.2019.1634196

Goldinger, S. D., & Papesh, M. H. (2012). Pupil Dilation Reflects the Creation and Retrieval of Memories. *Current Directions in Psychological Science*, *21*(2), 90-95. https://doi.org/10.1177/0963721412436811

Jacobs, H. I. L., Priovoulos, N., Poser, B. A., Pagen, L. H. G., Ivanov, D., Verhey, F. R. J., & Uludağ, K. (2020). Dynamic behavior of the locus coeruleus during arousal-related memory processing in a multi-modal 7T fMRI paradigm. *Elife*, *9*, e52059. https://doi.org/10.7554/eLife.52059

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*(1), 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*(10), 1971-1989. https://doi.org/10.1080/17470218.2011.588335

Kafkas, A. (2021). Encoding-linked pupil response is modulated by expected and unexpected novelty: Implications for memory formation and neurotransmission. *Neurobiology of Learning and Memory, 180*. https://doi.org/10.1016/j.nlm.2021.107412

Kafkas, A., & Montaldi, D. (2015). Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus, 25*(11), 1262-1273. https://doi.org/10.1002/hipo.22434

Kalwani, R. M., Joshi, S., & Gold, J. I. (2014). Phasic activation of individual neurons in the locus ceruleus/subceruleus complex of monkeys reflects rewarded decisions to go but not stop. *Journal of Neuroscience, 34*(41), 13656-13669. https://doi.org/10.1523/JNEUROSCI.2566-14.2014

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583-1585.

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE, 3*(1), e1532. https://doi.org/10.1371/journal.pone.0001532

Kleiner, M., Brainard, D., & Pelli, D.G. (2007). What's new in psychtoolbox-3? *Perception 36 ECVP Abstract Supplement.*

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., . . . Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, *8*(1), 4949. https://doi.org/10.1038/s41598-018-23197-6

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research, 24*(2), 326-334. https://doi.org/10.1016/j.cogbrainres.2005.02.005

Leow, L. A., Tresilian, J. R., Uchida, A., Koester, D., Spingler, T., Riek, S., & Marinovic, W. (2021). Acoustic stimulation increases implicit adaptation in sensorimotor adaptation. *European Journal of Neuroscience*, *54*(3), 5047-5062. https://doi.org/10.1111/ejn.15317

Li, J., & Deng, S. W. (2022). Facilitation and interference effects of the multisensory context on learning: a systematic review and meta-analysis. *Psychological Research*. https://doi.org/10.1007/s00426-022-01733-4

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PloS one, 13*(5), e0196391. https://doi.org/10.1371/journal.pone.0196391

Mathot, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. Behavior Research Methods, 50(1), 94-106. https://doi.org/10.3758/s13428-017-1007-2

Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *Elife*, *11*. https://doi.org/10.7554/eLife.70510

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248-251. https://doi.org/10.1126/science.287.5451.248

McGaugh, J. L. (2018). Emotional arousal regulation of memory consolidation. *Current Opinion in Behavioral Sciences*, *19*, 55-60. https://doi.org/https://doi.org/10.1016/j.cobeha.2017.10.003

Miller, A. L., & Unsworth, N. (2020). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(12), 2277–2294. https://doi.org/10.1037/xlm0000797

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision, 13*(11). https://doi.org/10.1167/13.2.11

Nassar, M. R., & Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., Gold, J. I. (2013). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience, 15*(7), 1040-1046. https://doi.org/10.1038/nn.3130

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods, 45*, 272–288. https://doi.org/10.3758/s13428-012-0247-4

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*, 1346-1353. https://doi.org/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. *International Journal of Psychophysiology, 83*(1), 56-64. https://doi.org/10.1016/j.ijpsycho.2011.10.002

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. Nature Reviews. *Neuroscience, 10*(3), 211-223. https://doi.org/10.1038/nrn2573

Sara, Susan J., & Bouret, S. (2012). Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. *Neuron, 76*(1), 130-141. https://doi.org/10.1016/j.neuron.2012.09.011

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6). 110-114.

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology, 16*(14), 1422-1427. https://doi.org/10.1016/j.cub.2006.05.048

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411-417. https://doi.org/10.1016/j.tics.2008.07.006

Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1962-1987. https://doi.org/10.1037/xlm0000421

Vo, M. L. H., Jacobs, A. M., Kuchinke, L., Hofman, 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. https://doi.org/10.1111/j.1469-8986.2007.00606.x

von Kriegstein, K. & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*(10), e326. https://doi.org/10.1371/journal.pbio.0040326

Wetzel, N., Einhäuser, W., & Widmann, A. (2020). Picture-evoked changes in pupil size predict learning success in children. *Journal of Experimental Child Psychology, 192*. https://doi.org/10.1016/j.jecp.2019.104787

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. NeuroImage, 39(2), 885-893. https://doi.org/https://doi.org/10.1016/j.neuroimage.2007.09.028

Young, S. G., & Hugenberg, K. (2012). Individuation motivation and face experience can operate jointly to produce the own-race bias. *Social Psychological and Personality Science, 3*(1), 80-87. https://doi.org/10.1177/1948550611409759

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends Hear*, *22*, 2331216518777174. https://doi.org/10.1177/2331216518777174