Journal of Experimental Psychology: Learning, Memory, and Cognition

Trial-Level Fluctuations in Pupil Dilation at Encoding Reflect Strength of Relational Binding

Jonathon Whitlock, Ryan Hubbard, Huiyu Ding, and Lili Sahakyan

Online First Publication, September 4, 2023. https://dx.doi.org/10.1037/xlm0001286

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001286

Trial-Level Fluctuations in Pupil Dilation at Encoding Reflect Strength of Relational Binding

Jonathon Whitlock[1], Ryan Hubbard[1, 2], Huiyu Ding[1], and Lili Sahakyan[1, 2]
[1] Department of Psychology, University of Illinois at Urbana-Champaign
[2] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign

Eye-tracking methodologies have revealed that eye movements and pupil dilations are influenced by our previous experiences. Dynamic fluctuations in pupil size during learning reflect in part the formation of memories for learned information, while viewing behavior during memory testing is influenced by memory retrieval and drawn to previously learned associations. However, no study to date has linked fluctuations in pupil dilation at encoding to the magnitude of viewing behavior at test. The current investigation involved monitoring eye movements both in single item recognition and relational recognition tasks. In the item task, all faces were presented with the same background scene and memory for faces was subsequently tested, whereas in the relational task each face was presented with its own unique background scene and memory for the face–scene association was subsequently tested. Pupil size changes during encoding predicted the magnitude of preferential viewing during test, as well as future recognition accuracy. These effects emerged only in the relational task, but not in the item task, and were replicated in an additional experiment in which stimulus luminance was more tightly controlled. A follow-up experiment and additional analyses ruled out differences in orienting instructions or number of fixations to the encoding display as explanations of the observed effects. The results shed light on the links between pupil dilation, memory encoding, and eye movement patterns during recognition and suggest that trial-level fluctuations in pupil dilation during encoding reflect relational binding of items to their context rather than general memory formation or strength.

Keywords: eye movements, pupil dilation, relational binding, preferential viewing, memory strength

Supplemental materials: https://doi.org/10.1037/xlm0001286.supp

Our previous experiences and memories of the past influence our behavior in many ways, including how we direct our eye movements and what stimuli in our environment we focus our gaze on. For instance, viewing behavior is affected by previous experience in that eye movements are drawn to previously learned associations (Hannula et al., 2007), and reflect the strength of relational binding of items to their studied context (Whitlock et al., 2020), even during instances in which conscious recollection is unsuccessful (Hannula et al., 2012; Hannula & Ranganath, 2009; Nickel et al., 2015; Whitlock et al., 2020). The use of eye-tracking methodology in memory research has largely focused on eye movements in response to visual stimuli (for a review, see Hannula et al., 2010) and fluctuations in pupil size (for a review, see Sirois & Brisson, 2014). Patterns of eye movements during learning that are reproduced at the time of test benefit retrieval of that learned information (Norton & Stark, 1971), in that the pattern of transitions and fixations within a visual display during both learning and test predicts the success of subsequent memory retrieval (Damiano & Walther, 2019; Johansson & Johansson, 2014; Richardson & Spivey, 2000). Studies utilizing pupillometry (i.e., the measurement of pupil size in response to presented stimuli) have revealed dynamic pupil size changes to be indicative of subsequent memory strength of learned information (Kafkas & Montaldi, 2011; Otero et al., 2011; van Rijn et al., 2012). Therefore, the manner in which viewing behavior is deployed and the physiological changes indexed by pupil fluctuations have been shown to significantly impact or reflect the creation of long-term memories. The current investigation was designed to combine these two eye-tracking methods and assess the extent to which pupil size fluctuations during encoding might predict viewing behavior at test. To our knowledge, this is the first study to examine the relationship between these two measures.

Task-Evoked Pupil Responses During Memory Encoding and Retrieval

Investigations into task-evoked pupillary reflexes (TEPRs) have traditionally focused on the extent to which they reflect cognitive effort devoted to a task (Hess & Polt, 1964; Hess, 1965; Hess et al., 1965), including as a function of task difficulty (Kahneman et al., 1967) and working memory demands (Heitz et al., 2008; Kahneman & Beatty, 1966). Pupil dilations were shown to index individual differences in sustained attention, both in the intensity and consistency of attention during word list learning (Unsworth & Miller, 2021a, 2021b) and

Jonathon Whitlock https://orcid.org/0000-0001-9815-5147
Ryan Hubbard https://orcid.org/0000-0001-6204-8729
The authors thank Skylar Mu, Emma Dawn, and Ivy Chen for their assistance with data collection. Data and analysis code is available online at the following link: https://osf.io/psf85/ (Whitlock et al., 2023).
Correspondence concerning this article should be addressed to Jonathon Whitlock, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820, United States. Email: jsw6@illinois.edu

studying of paired associates (Miller & Unsworth, 2020, 2021), as well as individual differences in working memory capacity (Unsworth & Robison, 2015). In long-term memory research, pupil size differs between correctly recognized old and new items during retrieval, a finding known as the *pupillary old/new effect* (Otero et al., 2011; Võ et al., 2008). The latter is thought to reflect the increased voluntary effort devoted to successful recognition of studied items (Goldinger & Papesh, 2012; Papesh et al., 2012; Võ et al., 2008; but see Mill et al., 2016). Additionally, memory strength differences of learned information are reflected in variations in pupil size during encoding as well as during retrieval (Kafkas & Montaldi, 2011; Otero et al., 2011; Papesh et al., 2012; van Rijn et al., 2012). However, the exact nature of pupil sizes during learning indexing successful encoding has been inconclusive.

In some cases, increased TEPRs during encoding predicted subsequent memory success, as in the hit rates (Ariel & Castel, 2014; Papesh et al., 2012), while others have concluded that larger TEPRs predicted subsequent forgetting, as in the miss rates (Kafkas & Montaldi, 2011; Naber et al., 2013). Still others found no relationship between TEPRs and subsequent memory, neither in hits, nor in the miss rates (Eldar et al., 2016; Võ et al., 2008; for a review, see Gross & Dobbins, 2021). In sum, while changes in pupil dilations at the time of study seem related to memory encoding, there is little consensus regarding the underlying memory processes that are reflected in TEPRs. Importantly, the focus of these previous studies was on single item memory such as individual words or images, with little to no consideration of relational memory demands. Item memory tests typically require discriminating between studied and nonstudied items and therefore can be supported by familiarity processes that indicate how likely it is for test items to have been recently encountered during learning (Gillund & Shiffrin, 1984; Hockley & Consoli, 1999; Hockley & Murdock, 1987). In contrast, relational memory involves arbitrary pairing of different elements of an event, such as names to faces associations, face-to-scene associations, or spatiotemporal relations (Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001). Tests of relational memory require discriminating between studied items that are presented either with their original relationship intact or a newly formed relationship, and therefore require recollection of the original learning episode. Therefore, different processes are believed to be engaged between item and relational memory tasks, and the latter may have its unique underlying process reflected by TERPs.

In the current study, complex images involving faces and scenes were studied in separate contexts that either emphasized or downplayed the relational component between face and scene. Task conditions that downplayed the relational component instead emphasized item information regarding the face. Memory for complex images involving faces and scenes require the binding of various elements contained in the image and thus require greater relational memory demands than tasks that only involve item information such as that for individual faces. Therefore, we expected greater relational binding processes to be engaged in conditions emphasizing encoding and retrieval of relational information regarding which faces were studied with test background scenes compared to conditions which emphasized only item information regarding which faces were studied. Given that subsequent memory effects become more pronounced as associative memory demands during learning increase (see Voss et al., 2017), subsequent memory effects during learning are likely to be more pronounced for task demands emphasizing relational information than task demands that downplay relational information and instead emphasize item information.

Previous investigations into TEPRs and memory outcomes have mostly examined the association of TEPRs with memory accuracy, which is a single and somewhat crude measure of memory, as it reflects a summary of an entire test trial within the confines of either successful or unsuccessful memory retrieval. In contrast to accuracy, viewing behavior has consistently been shown to be a very sensitive measure of memory retrieval (for a review, see Hannula et al., 2010), discussed in the section below. Thus, eye movements during retrieval may provide a more nuanced assessment of memory outcomes that are related to TEPRs during encoding, helping to disentangle some of the discrepant findings in the literature.

Viewing Behavior During Memory Encoding and Retrieval

Eye-tracking methodology allows assessing not only pupil size changes, but also complex pattern of viewing during memory encoding and retrieval. The manner in which viewing behavior is deployed during learning is associated with retrieval success as well as the quality of memory representations for learned information (Ferreira et al., 2008; Kafkas & Montaldi, 2011; Lucas et al., 2019; Wynn et al., 2018). Long-term memory for images improves with more fixations during encoding (Damiano & Walther, 2019; Kafkas & Montaldi, 2011; Loftus, 1972; Olejarczyk et al., 2014; Olsen et al., 2016). Additionally, regions of an image receiving higher fixation density are associated with greater rates of image recognition of those regions (van der Linde et al., 2009), and repeatedly sampling the same object is associated with improved memory for that object (Pertzov et al., 2009). Furthermore, allowing participants to freely view complex images such as scenes containing objects benefits subsequent memory retrieval compared to when viewing is restricted (Damiano & Walther, 2019; Henderson et al., 2005; Molitor et al., 2014). In relational memory, greater frequency of alternating viewing between items presented simultaneously is related to greater subsequent memory for which objects were studied together (Kamp & Zimmer, 2015). In sum, retrieval success and memory strength are associated with greater accumulation of visual information in a structured and organized fashion (Lucas et al., 2019).

Eye movements during retrieval can reflect memory for previously studied relations among items or items and the context in which they were encountered at learning, a finding known as the *preferential viewing effect* (Hannula et al., 2007). Preferential viewing demonstrates that eye movements reflect relational memory processes that are required for nuanced distinctions between equally familiar stimuli whose previously formed associations are the subject of the memory probe. These viewing effects emerge rapidly and obligatorily as early as 500–750 ms following the onset of the test display, and vary according to memory strength, emerging even when the subsequent behavioral response selection is inaccurate (Hannula et al., 2012; Hannula & Ranganath, 2009; Nickel et al., 2015; Whitlock et al., 2020). Therefore, preferential viewing can be a more sensitive marker of successful memory formation than the typical behavioral measures such as response accuracy. Since eye movements can reveal expressions of memory even in the absence of subsequent explicit behavioral endorsement of remembering, the viewing behavior at test might be an ideal candidate for assessing the relationship of TEPRs at encoding with downstream memory retrieval.

Paradigm for Studying Pupil-Size Fluctuations and Viewing Behavior

To assess TEPR profiles in item and relational memory, we modified a procedure used to evaluate eye movements relating to both item and relational memory in the same design (Baym et al., 2014) by including monitoring of pupil sizes during encoding. Specifically, this design allowed us to relate TEPRs at encoding to measures of preferential viewing at test. Participants studied faces against a background scene in both item and relational memory tasks while their eye movements were monitored. In the item task, each face was studied against the same background scene and testing involved simple item recognition, whereas in the relational task, each face was studied with its own unique background scene and testing involved recognizing which item was studied with the test background scene. We also provided orienting instructions during learning to emphasize face information in the item task and the relationship between face and scene in the relational task. We suspect differences in TEPR expressions of memory depend on how relational memory is emphasized during learning and probed for during retrieval. Furthermore, pupil dilations that reflect relational binding may only be related to viewing behavior during retrieval when those same processes are reengaged.

The benefit of combining the two types of memory tasks within the same paradigm is that it allows contrasting viewing behavior indicative of item memory from that of relational memory and the underlying mechanisms producing that behavior. For the present purposes, it also meant that pupillary responses during encoding between the two tasks could be compared within the same paradigm and thus can be contrasted between simple item encoding and more complex relational encoding. The critical question was whether the relationship between pupil fluctuations during learning and the magnitude of viewing test items would emerge equally in both the item and relational task, or alternatively that the different memory processes engaged by these two different tasks would lead to different relationships between pupil dilation at encoding and viewing behavior at test.

Method

Data

The pupil size and eye-movement data analyzed in this article was collected as part of a larger project on the cross-race effect (Ding et al., 2021). Note that there is no overlap in the hypotheses or the analyses reported in this article and the project from which this data were taken. Preliminary results including the effect of race in the following analyses revealed the effect of, and interactions involving, race were all not significant, all $ps > .202$. Therefore, the variable of race was not included in the following analyses.

Participants

Participants were 36 undergraduate students from the University of Illinois who participated in exchange for course credit. The study was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign and complied with American Psychological Association (APA) ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study. They were tested individually in the lab prior to COVID-19 pandemic.

Apparatus

Eye movements were recorded throughout the entire study at a rate of 1,000 Hz using an Eyelink 1000 Plus eye-tracking system (SR Research). Eye position was calibrated using a 3×3 spatial array before the beginning of each encoding and test phase, ending when participants fixated on a centrally located cross hair, indicating the beginning of each phase. The computer screen resolution was set to $1{,}280 \times 1{,}024$.

Stimuli

Stimuli consisted of 216 faces selected from Tullis et al. (2014) and the Face Research Lab London Set (DeBruine & Jones, 2017), presented in gray scale, and sized to 330×480 pixels. All faces depicted neutral expressions, displayed from the neck up. Faces consisted equally of male and female faces, and equally of Asian and Caucasian faces. Scene stimuli were selected from the Fine-Grained Image Memorability (FIGRIM) data set (Bylinskii et al., 2015). Background scene stimuli consisted of 109 colored scenes depicting outdoor environments such as rolling hills, beaches, and cities. None of the background scenes contain human faces. Face and scene pairings were randomly determined across participants. During encoding, faces were displayed in the center of the scene, whereas at test faces were displayed in the top left, top right, and bottom middle positions of the scenes. Each face was equally likely to be displayed in any of the three positions across participants, as well as had an equal chance of being a target or a lure during test.

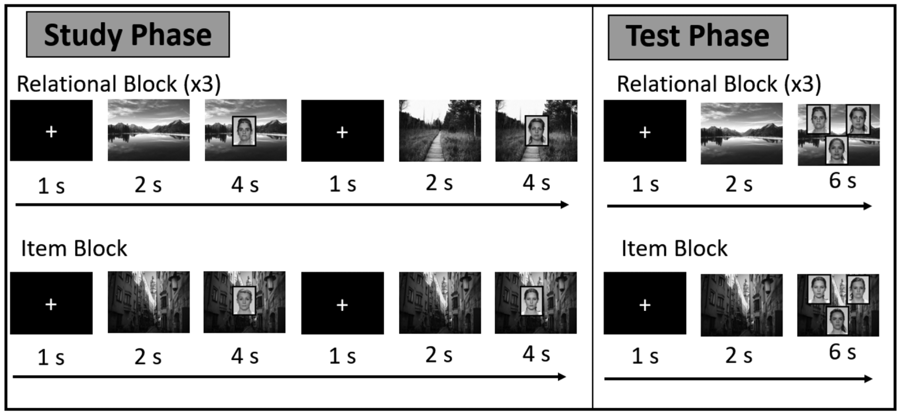
Procedure

The paradigm used in the experiment was based on the design used by Baym et al. (2014). The details of the procedure can be seen in Figure 1. Participants completed both item and relational tasks, while their eye movements and pupil sizes were monitored throughout the entire procedure. Both the item and relational task consisted of presenting faces superimposed on scenes during encoding. In the item task, each face was studied with the same scene, and participants were asked to think about whether they would be friends with that person, and therefore emphasized item information regarding the face (i.e., “friendliness instruction”). In the relational task, each face was studied with its own unique background scene, and participants were asked whether they could imagine that person in that background place, and therefore emphasized the relationship between face and scene (i.e., “integrate instruction”). The purpose of including the different orienting tasks across the two tasks was that each type of memory being tested was emphasized during encoding to ensure sufficient encoding of the appropriate information.

At study, a fixation cross was initially presented for 1 s, followed by an unobstructed scene for 2 s, and then a face was superimposed on the scene for 4 s. The test consisted of a three-face test display superimposed against a previously studied background scene. A fixation cross was initially presented for 1 s, followed by an

Figure 1

During Learning, Participants Are Presented With Face–Scene Pairs in Both an Item and Relational Task



Note. For the relational task, each face is studied with its own unique background scene, and at test, they select the face they remember studying with the background scene. For the item task, each face is studied with the same background scene, and at test, they select the face they remember studying. In the relational task, each lure face was studied on other background scenes and therefore is equally familiar, whereas in the item task, each lure face was not studied and therefore are novel faces.

unobstructed scene for 2 s, and then, a three-face test display was superimposed on the scene for 4 s, during which time the participants made their face selection response.

The relational task consisted of three study-test blocks. After studying 36 unique face–scene pairs, the test was administered, which involved 12 trials. Each test trial presented three previously studied faces against a previously studied background scene, where one of the faces had been studied with that scene and the other two faces were studied with other scenes, meaning three faces were equally familiar at the time of test. Participants' task was to indicate which face they remembered studying with that background scene. As is a common procedure with this paradigm, the number of trials in study and test phase in the relational task have a 3:1 ratio—namely, any given three study trials produce one test trial. For this reason, the relational task consisted of three separate study-test blocks, each involving 36 unique face–scene pairings during encoding, immediately followed by 12 test trials. The procedure was repeated three times with different stimuli to achieve the same number of test trials in the relational task (36) as in the item task (36), while also avoiding presenting too many study faces consecutively in one block during encoding in the relational task.[1]

The item task consisted of a single study-test block, wherein 36 face–scene pairs were studied against the same background scene. The study block was immediately followed by 36 test trials that involved presenting three faces against the background scene from encoding. Importantly, one of the faces was studied, whereas the other two faces were novel. Participants' task was to indicate which face they remembered studying during encoding. Note that the lures were completely new faces in the item task, whereas the lures in the relational task were all familiar because they were all studied with different scenes.

The race and gender of the faces in the test display were controlled so that all faces were either male or female, as well as either all Asian or all Caucasian. Thus, the faces in each test display were matched for both gender and race. The target face had an equal probability of being presented in the top left, top right, and bottom middle position, and faces were equally likely to be in the study or test phases. The order of the item and relational tasks was counterbalanced across participants.

A crucial difference between the relational and item tasks was that lures in the relational task were studied, albeit with their own unique background scene, whereas lures in the item task were nonstudied faces. This difference means pupil size fluctuations during learning were recorded for faces that subsequently went on to serve as lures during retrieval in the relational task only, affording us an opportunity to contrast pupil size fluctuations predicting viewing to faces presented both with their original as well as a different background scene during learning.

Analytic Approach

Preprocessing of Viewing Behavior

Viewing to faces at test was calculated as a proportion of time viewing a face out of total time viewing all three faces. To do so, we divided the amount of time viewing a face (t) from the total time viewing all faces (T; thus, calculated by t/T). Proportion of time viewing faces was aligned with respect to when the behavioral response was made (i.e., selecting one of three faces from the test display) on a trial-by-trial basis (i.e., a response-locked analysis). As is typical in studies employing this paradigm, we expected greater viewing to be devoted to selected targets compared to selected lures, denoting the preferential viewing effect. To analyze the emergence of preferential viewing, a mixed effects regression model was fit to the proportion of viewing on each trial, using *selected face*

[1] Pilot studies from our lab indicate that presenting 108 study faces in the relational task consecutively leads to a near chance performance at test.

(target vs. lure) and *bins* (four 500 ms time bins, equating to 2 s prior to response) as fixed effects, and *participants* as a random intercept.[2] This analyses is presented in Table 1. There was a significant effect of Selected Face, indicating a greater proportion of viewing devoted to the selected target than the selected lure prior to the behavioral response, replicating previous findings of preferential viewing. Moreover, consistent with published findings, the magnitude of preferential viewing varied across the time course, as evidenced by the significant Selected Face × Bins interaction (see Figure 2).

In order to relate pupil size during encoding to viewing behavior at test, we first identified where in the time course eye-movement expressions of memory were at their greatest. We operationalized this by identifying when the magnitude of preferential viewing peaked and stabilized (from now on referred to as the critical time window), and performed targeted analyses within this time Window. There was a significant Selected Face × Bins interaction between the time bins of −2,000 and 1,500 ms, as well as between the time bins of −1,500 and −1,000 ms, indicating that preferential viewing became larger across these time bins. The Selected Face × Bins interaction was not significant between the time bins of −1,000 and −500 ms, indicating that preferential viewing had peaked and stabilized within these time bins (see Figure 2). Therefore, the following analyses focus on the pupil size during learning predicting the magnitude of viewing to test items within the critical time window, comprising a full second of viewing prior to behavioral selection on a trial-by-trial basis. Proportion of viewing to all test items, both selected and unselected, is visualized in Figure S1 in the online supplemental materials.

We note that when participants failed to make a response on a given trial, this precluded aligning viewing behavior to behavioral response on those trials, and therefore trials that did not include a response were not included in the final analysis. Overall, an average of 2% of trials were removed from each participant's data.

Preprocessing of Pupil Dilation

Pupil diameter was measured during the experiment from the left eye with a sampling rate of 1,000 Hz. Following data collection, the data were preprocessed to remove eye blinks and trials containing artifacts. The onsets and offsets of blinks were identified using an automated algorithm based on noise in the pupillometry signal (Hershman et al., 2018). Identified blinks were removed, and the data were interpolated with shape-preserving piecewise cubic interpolation. Trial epochs of 100 ms prior to the onset of the scene to 6 s after the onset of the scene were then extracted from the pupillometry time-course, and the 100-ms prestimulus period was used to baseline correct each trial. Lastly, trials with large rapid deviations in dilation were automatically removed, and additional visual inspection was performed to remove trials with slower artifacts. Overall, an average of 8% of trials were removed from each subject's data, of which 75% of trials removed were in the relational task and the remaining 25% of removed trials were in the item task.[3] Trial-level pupil dilation measurements were obtained by extracting the mean of the signal from 2 to 4 s after the onset of the face stimulus to avoid influences of stimulus luminance as well as target the window where dilation effects are often observed (Gross & Dobbins, 2021; Mill et al., 2016). Pupil dilation values were *z*-scored for plotting purposes for easier comparison across tasks.

Table 1

Outcomes of Mixed Effects Model Regression Analysis of Viewing to the Selected Target and Selected Lure Faces Collapsed Across the Item and Relational Task, Using Selected Face (Target vs. Lure) and Bins as Fixed Effects and a Random Intercept for Participants (Experiment 1)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Bins	0.15	0.01	16.84	<.001**
Selected face	0.06	0.01	6.13	<.001**
Bins × Selected Face	0.10	0.02	5.63	<.001**
Follow-up analysis to Bins × Face interaction				
Bins −2,000 and −1,500				
Bins × Selected Face	0.06	0.03	2.14	.032*
Bins 1,500 and −1,000				
Bins × Selected Face	0.09	0.02	3.55	<.001**
Bins 1,000 and −500				
Bins × Selected Face	−0.02	0.02	0.78	.438

* $p < .05$. ** $p < .001$.

Viewing Time and Pupil Dilation Analyses

We assessed whether pupil size during encoding predicted the proportion of time viewing the target face at test as a function of whether it was subsequently selected, and how this relationship varied between the item and relational tasks. Importantly, the variable of selection specifically refers to whether a test item in the analysis would go on to be subsequently selected on that trial. Assessment of proportion of viewing time was limited to the critical time window comprising 1 s prior to behavioral response. Analyses were performed with linear mixed effects regression models in R (Bates et al., 2014), which allowed for analyses at the trial level, as well as accounting for participant-level variance. A linear mixed effects model was initially fit to predict viewing behavior (the magnitude of viewing targets on each trial) for both item and relational tasks, including *pupil* (pupil size fluctuations as a continuous measure), *selection* (selected vs. unselected), and *task* (item vs. relational) as fixed effects and a random intercept for *participants*.

Note, that the effect of selection on the outcome measure of viewing would reflect greater viewing for selected than unselected items, whereas an interaction involving selection would indicate that the subsequent selection of items would depend on the effect of another interacting variable. Given that previous research demonstrates greater viewing to selected than unselected items, we expected a main effect of selection to be present in all analyses. However, the interaction of task and pupil with selection was of main interest in the following analyses involving viewing time and pupil dilation.

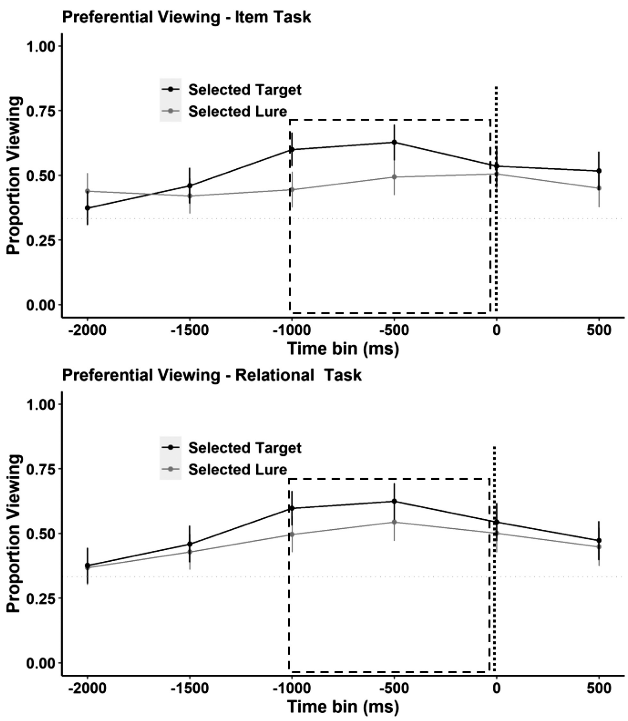
Given that lures in the relational task were presented during encoding when pupil size recording occurred, we were able to conduct a similar analysis involving the lures in the relational task. Therefore, a second mixed effects model was fit to predict viewing behavior in the relational task only, including pupil dilation and

[2] Due to the inclusion of random slopes in the models resulting in singular fits to the data, random slopes were not included in any analysis involving viewing and pupil size. However, random slopes were included in models that assessed recognition accuracy, as these models did not result in singular fits to the data.

[3] A similar ratio of trials was removed from the data in Experiment 3 as in Experiment 1. Specifically, 75% of removed trials were from the relational task, wherein the remaining 25% of trials were from the item task.

Figure 2

Proportion of Viewing to the Selected Target and Selected Lure in Experiment 1 Across Time Bins (Separated by 500 ms) During Testing for Each Task, Collapsed Across Race, and Aligned to the Behavioral Selection on a Trial-by-Trial Basis (Response-Locked Analysis)



Note. Vertical dotted line represents the point in time in which participants made their behavioral selection, and the dashed box indicates the critical time window.

selection (selected vs. unselected) as fixed effects and a random intercept for participants. Critically, the selection variable refers to whether the lure was selected on that trial.

We also conducted a follow-up analysis aimed at identifying between-task differences in the number of fixations deployed during learning and how this relates to the pupil size on a trial-by-trial basis. The two tasks differed in two important respects, namely, the nature of the background scene during learning as well as the orienting instructions provided to participants during learning. Specifically, providing unique background scenes for each face in the relational task, as well as instructions meant to encourage binding of item and context, contrasted with the item task in which the background scene was the same on each trial and the orienting instructions encouraged only item-level information. The net result of these differences was likely to be a greater number of fixations to the visual display in the relational than the item task. Thus, this analysis assesses whether there are differences in the number of fixations between the item and relational task and whether this affects the pupil size on a trial-by-trial basis. A linear mixed effects model was fit to the pupil size data for both item and relational tasks, using (number of) fixations and task as fixed effects and a random intercept for participants. Data and analysis code is available online at the following link: https://osf.io/psf85/.

Results

Recognition Accuracy

We first examined recognition accuracy between the item and relational tasks. A mixed effect logistic regression model was fit to recognition accuracy on each trial, using *task* (item vs. relational) as a main effect, *participants* as a random intercept, and random slopes for *task*.[4] The fixed effect of task was significant, $\beta = .72$, $SE = 0.15$, $Z = 4.84$, $p < .001$, indicating that there was greater accuracy in the item task ($M = 0.74$, $SD = 0.49$) than the relational task ($M = 0.61$, $SD = 0.44$), replicating previous findings using this same procedure (Baym et al., 2014).

Pupil Size Fluctuations Predict Viewing to Selected Targets in Relational but Not Item Memory

We assessed whether fluctuations in pupil size at encoding could predict the magnitude of viewing to targets in the test display across item and relational tasks. A mixed effects regression model was fit to the proportion of viewing to targets on each trial, using *pupil size*, *selection* (selected vs. unselected), and *task* (item vs. relational) as fixed effects, and a random intercept for *participants*. There was a significant effect of selection, indicating more viewing was devoted to targets that were selected compared to targets that were not selected. There was also a significant effect of task, indicating greater viewing to targets in the item than the relational task (collapsed across accuracy). There was also a Task × Selection interaction, indicating greater difference in viewing selected than unselected targets in the relational task (approximately 30% difference) than in the item task (approximately 15% difference). Critical to our central interests, there was a significant Pupil × Selection × Task interaction (see Figure 3). To follow up the three-way interaction, we split the subsequent analyses based on the task variable (for details, see Table 2).

Specifically, in the item task, neither the main effect of pupil, nor the Pupil × Selection interaction were significant. In contrast, in the relational task, the main effect of pupil was not significant, but the Pupil × Selection interaction was significant. This was because of the significant effect of pupil for selected targets, but not for unselected targets. Thus, for targets that would go on to be subsequently selected during retrieval, greater pupil size during encoding was related to a greater magnitude of viewing those targets in the full second prior to their selection. Importantly, this relationship was not observed for targets that would not go on to be subsequently selected.

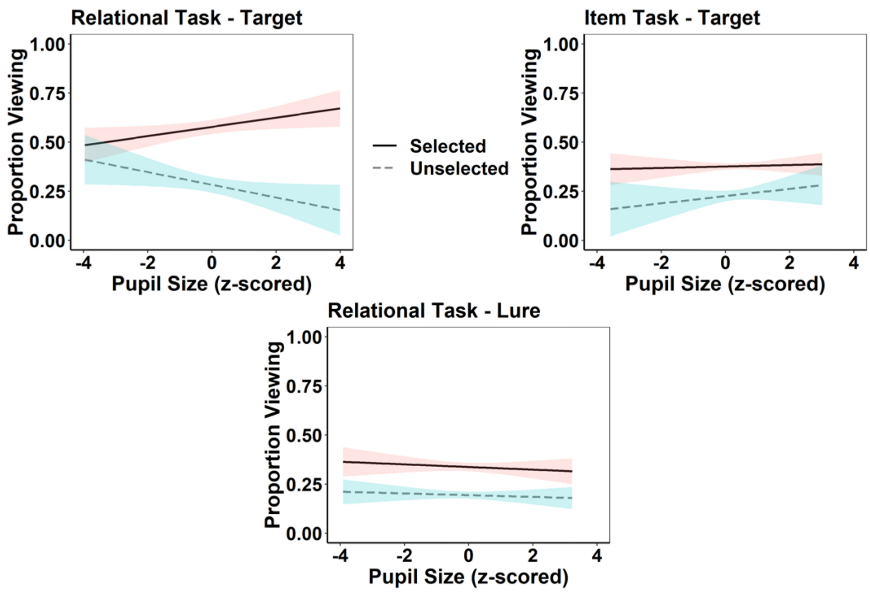
Pupil Size Fluctuations Do Not Predict Viewing to Selected Lures in Relational Memory

As described previously, pupil size fluctuations during learning were recorded for faces that subsequently went on to serve as lures in the relational task. Therefore, we performed an analysis limited to the relational task to assess if pupil size fluctuations predict

[4] We tested whether random slopes for the effect of task would contribute significantly to our mixed effect logistic regression model. Doing so revealed the model was improved by including a random slope for the effect of task, mixture $\chi^2_{2,1} = 26.87$, $p < .001$, and therefore random slopes were included for the effect of task in the final model.

Figure 3

Proportion of Viewing to Target Faces in the Relational and Item Task, and Lure Faces in the Relational Task, in the 1-s Time Bin Prior to Response, Across Standardized Pupil Size Differences During Encoding, in Experiment 1



Note. Greater values for pupil indicate greater pupil size at encoding. The units of pupil dilation, shown on the *x*-axis, reflect *z*-scored values of the raw pupil size area. Error bands reflect 95% confidence intervals. See the online article for the color version of the figure.

viewing to any selected items more generally, regardless of whether that item was paired with that background scene or not. Since lures were paired with different scenes at encoding and are therefore lacking the associative information with the presented scene, pupil size may be predicting a selection effect, rather than memory for relational information per se. To assess this possibility, a mixed effects regression model was fit to the proportion of viewing to lure faces on each trial, using *pupil size* and *selection* (selected vs. unselected) as fixed effects, and *participants* as a random intercept. The results are reported in Table 3 and visualized in Figure 3. Neither the main effect of pupil, nor the Pupil × Selection interaction were significant. Therefore, pupil size to lure faces in the relational task did not predict viewing to lure faces regardless of whether they were selected. This suggests that pupil size predicting viewing does not generalize to selected lures, indicating that pupil predicting viewing in the relational task is not reflecting a general effect of selecting a face, but instead reflects previous memory of seeing the target face paired with the background scene.

Number of Fixations to Visual Display Does Not Affect Pupil Size Between Tasks

To rule out the possibility that different number of fixations deployed to the visual display between the item and relational task may have impacted overall pupil size and thus could explain the previous results, we conducted the following analysis relating the number of fixations to the pupil size during learning on a trial-by-trial basis. A mixed effects regression model fit to the number of fixations made during each trial as a function of task (item vs. relational) revealed that there were greater number of fixations made in the relational than the item task. These results, and well as those of the following analysis, are reported in Table 4.

Given the greater number of fixations deployed to the visual display in the relational ($M = 3.2$, $SD = 2.14$) than the item task ($M = 1.48$, $SD = 1.09$), we assessed whether pupil size differed as a function of task and the number of fixations made to the display on each trial. A mixed effects regression model was fit to the average pupil size on each trial, using (number of) *fixations* and *task* (item vs. relational) as fixed effects and a random intercept for *participants*. There was a significant effect of task, indicating that pupil sizes were larger in the item task ($M = 0.28$, $SD = 0.74$) than the relational task ($M = 0.10$, $SD = 1.06$). The results are shown in Figure 4. There was no effect of fixation, indicating that the number of fixations on each trial did not affect pupil size. The Fixation × Task interaction was also not significant, indicating that the relationship between number of fixations and pupil size did not differ between the item and relational task. Thus, while the number of fixations differed between the two tasks, this was not predictive of pupil size differences and therefore was unlikely to explain the previous effects.

Discussion

In Experiment 1 we assessed whether pupil size fluctuations during learning predicted the magnitude of viewing test items prior to their selection. If pupil size fluctuations during learning predicted the magnitude of time spent viewing that item during retrieval, then they could potentially reveal important aspects of the physiological basis of encoding processes. These encoding processes could be general principles of memory formation or they could be

Table 2

Outcomes of Mixed Effects Model Regression Analysis of Viewing to the Target Faces in the Item and Relational Task, Using Pupil, Task (Item vs. Relational), and Selection (Whether the Test Item in the Analysis Would Go on to Be Subsequently Selected) as Fixed Effects and a Random Intercept for Participants (Experiment 1)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Pupil	<0.01	0.01	0.56	.576
Task	−0.13	0.01	9.99	<.001**
Selection	0.22	0.01	17.00	<.001**
Pupil × Task	0.01	0.02	0.96	.339
Pupil × Selection	0.02	0.02	1.48	.139
Task × Selection	−0.15	0.03	5.79	<.001**
Pupil × Task × Selection	−0.06	0.03	2.04	.042*
Follow-up analyses to three-way interaction, split by task variable				
Item task				
Pupil	0.01	0.01	1.06	.289
Selection	0.15	0.02	9.27	<.001**
Pupil × Selection	−0.01	0.02	0.70	.483
Relational task				
Pupil	−0.00	0.01	0.46	.645
Selection	0.29	0.02	15.07	<.001**
Pupil × Selection	0.06	0.02	3.06	.002*
Follow-up analysis to two-way interaction in relational task				
Selected target	0.02	0.01	2.29	.022*
Unselected target	−0.03	0.02	1.70	.089

* $p < .05$. ** $p < .001$.

unique to a specific kind of memory formation and thus may differ between item and relational recognition memory.

Pupil size fluctuations during the learning period predicted the magnitude of viewing correctly selected faces during the test period in the relational task. In contrast, pupil size fluctuations during learning did not predict the magnitude of viewing test items in the item task regardless of whether they were studied or nonstudied faces, and also regardless of whether they would subsequently be selected. Pupil size fluctuations during learning also failed to predict the magnitude of viewing incorrectly selected test items in the relational task.

Critical differences between the item and relational task presumably played a role in the divergent findings between the two tasks. Faces in the item task were all studied on the same background scene during learning, and the recognition test involved discriminating between studied and nonstudied faces. Importantly, the background scene was not critical to discriminating between studied and nonstudied faces (although we do not rule out implicit effects of reinstating the learning context during retrieval). In contrast, the background scene in the relational task was unique to each studied face and was critical in recognizing the face that it was originally studied with during retrieval. Previous research involving this task showed viewing behavior at test to be sensitive to the relationship between items and their originally studied context (Hannula et al., 2007; Hannula & Ranganath, 2009), and that memory strength for correctly recognized items in the relational task reflects the strength of successfully recognized items (Whitlock et al., 2020). Therefore, pupil size fluctuations during learning presumably predict the magnitude of relational binding of faces and their studied background scenes, and this is revealed when memory is probed for the previously formed relationship between face and background scene. More targeted analyses in the relational task contrasting target and lure faces ruled out the possibility that this relationship between pupil size at encoding and viewing at test was for any face that was subsequently selected. Both of these findings suggest that pupil size fluctuations during learning are more indicative of the success of relational binding rather than general memory formation processes.

An important issue with the current results is that the overall pupil dilations between the two tasks differed significantly. We note two critical differences between the item and relational task—namely, the orienting task participants were given during encoding differed, as well as whether the faces during learning were presented on the same compared to their own unique background scene. Instructing participants to assess the relationship between face and scene in the relational task could have elicited a greater number of fixations across the display compared to when they are instructed to assess the "friendliness" of the face, and potentially could have led to greater visual imagery in the relational task, which could have influenced pupil dilations (Miller & Unsworth, 2020). We conducted analyses to rule out the impact of fixations on pupil dilation, but could not rule out the potential differences because of visual imagery or nature of encoding. Furthermore, presenting different scenes in the background only during the relational task could have introduced a luminance confound—namely, the different scenes could emit different amounts of luminance, and this could lead to greater pupil dilation variability in the relational task compared to the item task. Given that the "friendliness" instruction was always for the item task in which the same background scene was used for all faces, both of which are likely to result in reduced viewing of the background scene as well as fewer fixations across the display in total, we conducted a follow-up study to assess the impact of these manipulations on pupil size, in order to rule out the possibility that orienting instructions and scene manipulation could explain differences in pupil size fluctuations between the item and relational task.

Table 3

Outcomes of Mixed Effects Model Regression Analysis of Viewing to the Lure Face in the Relational Task, Using Pupil and Selection (Whether the Test Item in the Analysis Would Go on to Be Subsequently Selected) as Fixed Effects and a Random Intercept for Participants (Experiment 1)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Pupil	−0.01	0.01	0.86	.388
Selection	−0.14	0.01	11.03	<.001**
Pupil × Selection	<0.01	0.01	0.18	.854

** $p < .001$.

Experiment 2

The aim of Experiment 2 was to assess the potential impact that orienting instructions and presenting the same compared to trial-unique background scenes on pupil size fluctuations on a trial-by-trial basis. Specifically, we were interested in whether the orienting instructions, scene manipulation, or both, resulted in differences in pupil size between the two tasks. Given that both varied simultaneously between the two tasks, we created four orthogonal conditions to tease apart the impact of orienting instructions and scene manipulation.

Table 4
Outcomes of Mixed Effects Model Regression Analysis of Number of Fixations Relating to Pupil Size in Both the Item and Relational Tasks, Using (Number of) Fixations and Task (Item vs. Relational) as Fixed Effects and a Random Intercept for Participants (Experiment 1)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Fixation	0.01	0.01	0.75	.454
Task	0.28	0.05	5.41	<.001**
Fixation × Task	0.05	0.03	1.92	.055
	Number of fixations differ by task			
Task	−1.71	0.06	29.66	<.001**

** $p < .001$.

Method

Participants

Participants were 43 undergraduate students from the University of Illinois who participated in exchange for course credit. The study was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign and complied with APA ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study. They were tested individually in the lab two years into the COVID-19 pandemic. The study complied with University-mandated safety protocols.

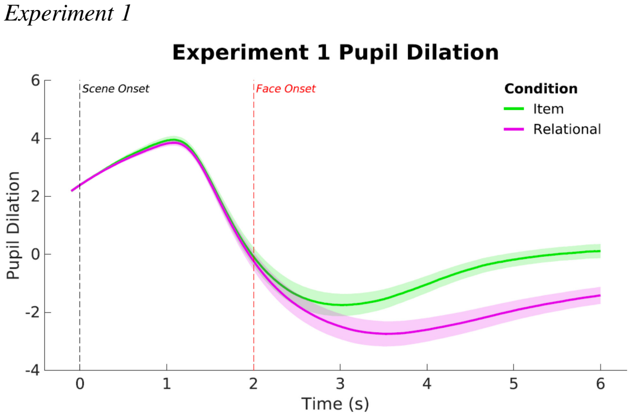
Apparatus

The apparatus, calibration procedures, and screen resolution was identical to those used in Experiment 1.

Stimuli

Stimuli consisted of 144 faces selected from a database of nonfamous people used in previous research (Althoff & Cohen, 1999). Care was taken to match for clothing, hairstyles, and professional lighting. Faces were cropped using the editing tool Adobe Photoshop so that everything above the chin was visible. Face images were sized to 300 × 300 pixels and were presented in color. All faces depicted neutral expressions, displayed from the neck up. Faces consisted equally of male and female faces. Importantly, since there was no effect of race in Experiment 1, we included only Caucasian faces. Scene stimuli were identical to those used in Experiment 1 and were selected from the FIGRIM data set (Bylinskii et al., 2015), and consisted of 74 unique scene images. Face and scene pairings were randomly determined across participants.

Figure 4
Pupil Size Fluctuations Across the Full 6 s Time Course of Experiment 1



Note. An unobstructed scene was presented for 2 s followed by a face superimposed on that scene for 4 s. Greater values of pupil dilation indicate larger pupil sizes. The units of pupil dilation, shown on the *y*-axis, reflect *z*-scored values of the raw pupil size area. Error bands reflect standard error around the mean. See the online article for the color version of the figure.

Procedure

Faces were presented superimposed on background scenes in a similar fashion as the encoding conditions of Experiment 1, with the exception that memory for faces was not subsequently tested, and participants were not expecting a memory test. Participants completed four separate blocks in a 2 (orienting instruction: friendliness vs. integrate) × 2 (scene: same vs. different) within-subject design. The friendliness instruction involved participants deciding whether they would be friends with that person, similar to the item task in Experiment 1, whereas the Integrate instruction involved deciding whether they could imagine that person in that background scene, similar to the relational task in Experiment 1. Additionally, for each instruction, half of the faces were presented with the same background scene and the other half of the faces were presented with different background scenes, in separate blocks. Blocks were randomized across participants so that half of participants first received the friendliness instruction whereas the other half first received the Integrate instruction. For all participants, the block containing same scenes across trials was presented before the block containing different scenes across trials, for both the friendliness and the integrate instructions.

During each trial, a fixation cross was initially presented for 1 s, followed by an unobstructed scene for 2 s, and finally a face superimposed on that scene for 4 s. Each block included 36 face–scene pairs, for a total of 144 trials.

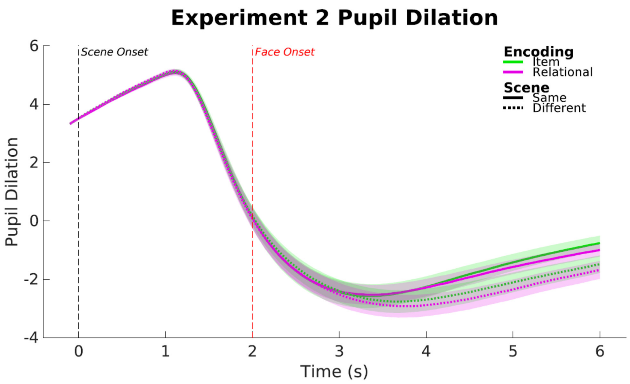
Results

Pupil Size Varies as a Function of Same Versus Different Scene Presentation

Pupil sizes varied across the time course for both item and relational tasks, as shown in Figure 5. Pupil sizes were extracted from the time window of 4–6 s (i.e., the last 2 s) of each trial, the mean of which was used for statistical analyses. Preprocessing of pupil dilation data were identical to the procedure used in Experiment 1. Analyses were performed with linear mixed effects regression models in R (Bates et al., 2014).

A mixed effects regression model was conducted, using *scene* (same every trial vs. different every trial) and *instruction* (friendliness vs. integrate) as fixed effects and a random intercept for *participants*. The results are reported in Table 5. There was a significant effect of scene, indicating that pupil sizes were overall larger when faces were presented on the same background scene compared to when they were presented on their own unique background scene. The effect of Instruction was not significant, indicating that pupil

Figure 5
Pupil Size Fluctuations Across the Full 6 s Time Course of Experiment 2



Note. An unobstructed scene was presented for 2 s followed by a face superimposed on that scene for 4 s. Greater values of pupil dilation indicate larger pupil sizes. The units of pupil dilation, shown on the *y*-axis, reflect *z*-scored values of the raw pupil size area. Error bands reflect standard error around the mean. See the online article for the color version of the figure.

sizes did not differ between the two orienting instructions. Finally, the Scene × Instruction interaction was not significant.

Discussion

The purpose of Experiment 2 was to identify possible confounding effects of orienting task and the scene manipulation varying whether faces were presented with the same or different background scenes as other faces. Pupil sizes were larger when faces were presented on the same background scene compared to when they were presented on their own unique background scene. In contrast, the orienting instructions provided to participants during face presentation did not have any impact on pupil size. Therefore, the larger pupil sizes overall in the item than the relational task in Experiment 1 was likely because of faces being presented on the same compared to different background scenes as other faces. The purpose of the following experiment was to replicate the findings of Experiment 1 with some modifications to control for differences in luminance across the two tasks.

Experiment 3

The purpose of Experiment 3 was to replicate the main findings of Experiment 1 with a larger sample size, and introduce a few minor modifications aimed at minimizing any differences in overall pupil size between the item and relational tasks. Specifically, in order to control for luminance effects when transitioning from a black screen with a centered white fixation point to a scene colored image, we used scrambled versions of the background scenes that maintained the luminance properties of the scenes. These "luminance masks" were presented immediately prior to presenting the background scene by itself (i.e., the "scene preview"), during both encoding and retrieval phases. This allowed the pupil size to become adjusted to the degree of brightness of the scenes before presenting the actual scene, possibly mitigating any differences in luminance between the two tasks. Additionally, in order to center participants' viewing prior to onset of the test display, a white fixation point against a black background was presented following the scene preview and prior to the onset of the test display with faces. This was done to encourage centering of fixations to the display, to decrease the chances that participants view a region of the scene containing faces before those faces are presented.

We also address the issue of TEPRs predicting accuracy as is typical in studies assessing TEPR reflections of memory. Experiment 1 established the relationship between TEPRs during learning and the viewing measure at test, indicating that the manner in which test items are viewed is critical in identifying TEPR expressions of memory. Therefore, in addition to replicating the original results from Experiment 1, we aim to examine whether pupil sizes during learning predict subsequent memory accuracy while taking into account the manner in which test items are viewed prior to their selection.

Table 5
Outcomes of Mixed Effects Model Regression Analysis of Pupil Size to Faces, Using Scene and Instruction as Fixed Effects and a Random Intercept for Participants (Experiment 2)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Scene	−0.19	0.02	8.00	<.001**
Instruction	0.04	0.02	1.45	.147
Scene × Instruction	0.02	0.05	0.50	.618

** $p < .001$.

Method

Participants

Participants were 76 undergraduate students from the University of Illinois who participated in exchange for course credit. The study was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign and complied with APA ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study. They were tested individually in the lab 2 years into the COVID-19 pandemic. The study complied with University-mandated safety protocols.

Apparatus

The apparatus, calibration procedures, and screen resolution was identical to Experiment 1.

Stimuli

Stimuli consisted of the 144 faces selected from the same set of faces used in Experiment 2 as well as an additional 72 faces for a total of 216 faces. The same criteria used to select faces in Experiment 2 were used for Experiment 3. There were 109 background scenes selected from those used in Experiments 1 and 2, specifically 108 background scenes for the relational task and one for the item task. During encoding, faces were displayed in the center of the scene, whereas at test faces were displayed in the top left, top right, and bottom middle positions of the scenes. Each face was equally likely to be displayed in any of the three positions across participants, as well as had an equal chance of being a target or a lure during test.

The luminance masks were created by scrambling the background scene images to remove any scene information while retaining low-level visual properties of the image (Grill-Spector et al., 1998). Specifically, box scrambling was used, in which each scene image was divided into nonoverlapping 5 × 5 blocks of pixels, and the pixel blocks were then randomly shuffled. An example of a scene and its shuffled counterpart are displayed in Figure S2 in the online supplemental materials. We confirmed the equivalent luminance between the shuffled and nonshuffled images using the SHINE Toolbox (Willenbockel et al., 2010).

Procedure

The procedure to Experiment 3 was similar to that of Experiment 1 with a few minor exceptions. For both study and test phases, a luminance mask was presented for 1.5 s following the fixation cross and prior to the scene preview in order to control for changes in luminance associated with transitioning from a black scene to a colored image with different luminance properties. Additionally, during the test phase, there was a 0.5-s centrally presented fixation cross that separates the scene preview and the onset of the test display. This was done in order to focus participants' attention to the center of the screen prior to presenting the test display.

Results

Recognition Accuracy

The first analysis examined whether there were recognition accuracy differences between the item and relational task. A mixed effect logistic regression model was fit to recognition accuracy on each trial, using *task* (item vs. relational) as a fixed effect, *participants* as a random intercept, and random slopes for *task*.[5] The fixed effect of task was significant, $\beta = .60$, $SE = 0.10$, $Z = 5.82$, $p < .001$, indicating that there was greater accuracy in the item task ($M = 0.72$, $SD = 0.45$), than the relational task ($M = 0.60$, $SD = 0.49$), replicating the findings of Experiment 1.

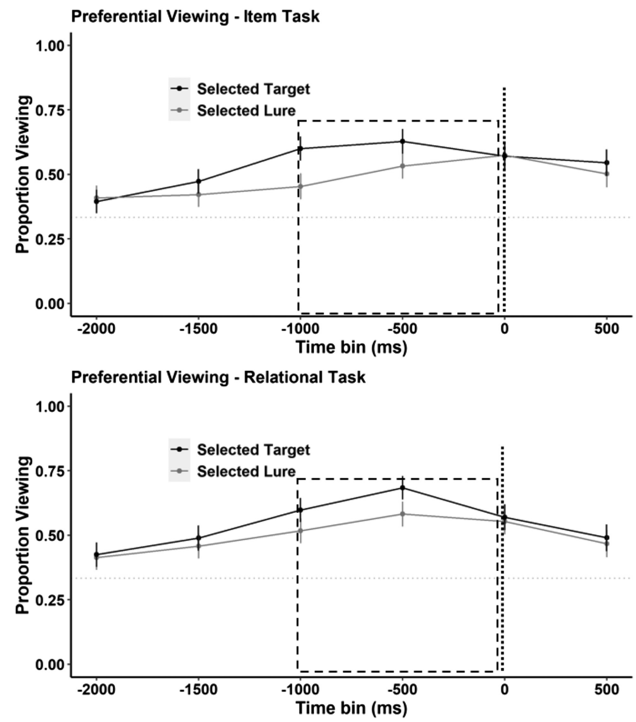
Pupil Size Predicts Viewing to Both Selected and Unselected Targets in Relational Memory but Not Item Memory

Item and Relational Task. For replication purposes, the same critical time window was used for analyses as in Experiment 1, visualized in Figure 6. Proportion of viewing to all test items, both selected and unselected, are visualized in Figure S3 in the online supplemental materials. For central analyses aimed at replicating Experiment 1 findings, a mixed effects regression model was fit to the proportion of viewing to targets on each trial, using *pupil size, selection* (selected vs. unselected), and *task* (item vs. relational) as fixed effects and a random intercept for *participants*. The results are reported in Table 6.

There was a significant effect of selection, indicating more viewing was devoted to selected compared to unselected targets. There was a significant Task × Selection interaction, indicating a greater difference in viewing selected targets compared to unselected targets in the relational task (approximately 46% difference) than in the item task (approximately 38% difference).

Figure 6

Proportion of Viewing to the Selected Target and Selected Lure Across Time Bins (Separated by 500 ms) During Testing for Each Task, and Aligned to the Behavioral Selection on a Trial-by-Trial Basis (Response-Locked Analysis) in Experiment 3



Note. Vertical dotted line represents the point in time in which participants made their behavioral selection, and the dashed box indicates the critical time window. Error bands reflect standard error of the mean.

Critical to the goals of our replication study, there was a significant Pupil × Selection × Task interaction (see Figure 7). Following up this interaction consisted of analyzing the fixed effects of selection and pupil size separately for each task. Doing so revealed a significant Pupil Size × Selection interaction in the relational task, but not in the item task. In turn, this two-way interaction in the relational task was because of pupil size significantly predicting the magnitude of viewing for both selected targets and unselected targets but in opposite directions, namely, a significant positive slope for selected targets and a significant negative slope for unselected targets. Therefore, the three-way interaction is because of greater pupil dilation during encoding predicting greater viewing for targets as a function of whether they would be selected in the relational task only, whereas no effect of pupil size was observed for targets in the item task regardless of whether they were selected.

[5] We tested whether random slopes for the effect of task would contribute significantly to our mixed effect logistic regression model. Doing so revealed the model was improved by including a random slope for the effect of task, mixture $\chi^2_{2,1} = 54.16$, $p < .001$, and therefore random slopes were included for the effect of task in the final model.

Table 6

Outcomes of Mixed Effects Model Regression Analysis of Viewing to the Target Faces in the Item and Relational Task, Using Pupil, Task (Item vs. Relational), and Selection (Whether the Test Item in the Analysis Would Go on to Be Subsequently Selected) as Fixed Effects and a Random Intercept for Participants (Experiment 3)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Pupil	<−0.01	0.01	0.97	.334
Task	<0.01	0.01	0.18	.859
Selection	0.36	0.01	35.72	<.001**
Pupil × Task	−0.01	0.01	1.37	.171
Pupil × Selection	0.02	0.01	2.47	.013*
Task × Selection	−0.03	0.02	1.37	.172
Pupil × Task × Selection	−0.06	0.02	2.82	.005*
Follow-up analyses to three-way interaction, split by task variable				
Item task				
Pupil	−0.01	0.01	1.71	.097
Selection	0.35	0.02	22.90	<.001**
Pupil × Selection	<−0.01	0.01	0.24	.813
Relational task				
Pupil	<0.01	0.01	0.52	.836
Selection	0.38	0.01	27.89	<.001**
Pupil × Selection	0.05	0.01	3.88	<.001**
Follow-up analysis to two-way interaction in relational task				
Selected target	0.03	0.01	2.78	.006*
Unselected target	−0.02	0.01	2.28	.023*

* $p < .05$. ** $p < .001$.

Pupil Size Does Not Predict Viewing to Selected Lures in Relational Memory

Relational Task Lures. Similar to Experiment 1, we were interested in conducting a targeted analysis within the relational task involving lure faces. The purpose of this analysis was to assess whether pupil size predicted viewing to targets generalized to any face studied within the relational task or if it was only the case with target faces (i.e., faces that were originally studied with the background test scene). A mixed effects regression model was fit to the proportion of viewing to lure faces on each trial, using *pupil* size and *selection* (selected vs. unselected) as fixed effects, and *participants* as a random intercept. The results are reported in Table 7 and visualized in Figure 7. Neither the main effect of pupil size nor the Pupil Size × Selection interaction was significant. Therefore, similar to Experiment 1, the finding of pupil size fluctuations predicting the magnitude of viewing test items in the relational task does not generalize to all faces studied with unique background scenes while memory is probed for the relation between test items and the background scene; instead, it is only specific to target faces, thereby ruling out these results being a selection effect.

Pupil Size at Encoding Predicts Accuracy for Relational Memory, but Not Item Memory

Given that we controlled for luminance and had greater power in Experiment 3, we included an additional exploratory analysis to examine the relationship between pupil size at encoding and behavioral accuracy at test. A mixed effects logistic regression model was fit to recognition accuracy on each trial, including *pupil size*, *viewing to target faces* (in the critical time window[6]), and *task* (item vs. relational) as fixed effects and a random intercept for *participants*. The results are reported in Table 8. The fixed effect of task was significant, as reported earlier in the results section on recognition accuracy. The fixed effect of viewing was also significant, indicating that there was a greater magnitude of viewing when participants would subsequently make a correct compared to an incorrect response. There was also a significant Pupil Size × Task interaction, indicating the relationship between pupil size and recognition memory accuracy differed between the two tasks. There was also a significant Task × Viewing interaction, indicating the relationship between viewing behavior and recognition memory accuracy differed between the two tasks. Finally, there was a significant Task × Pupil Size × Viewing interaction, because of a significant Pupil Size × Viewing interaction in the relational task, but not the item task (see Figure 8). Given that both variables were continuous variables, to unpack the interaction term in the relational task, we conducted simple slopes analysis by examining the impact of pupil size on accuracy at low (−1 *SD*) levels of viewing, mean (0 *SD*) level of viewing, and high (+1 *SD*) levels of viewing. This analysis showed that increased pupil size at encoding predicted greater memory accuracy when viewing was high. Pupil size did not predict memory accuracy at intermediate levels of viewing. Finally, at low levels of viewing, increased pupil size at encoding predicted lower memory accuracy. Thus, the extent to which pupil size at encoding predicted relational memory accuracy was contingent on the levels of viewing at test.

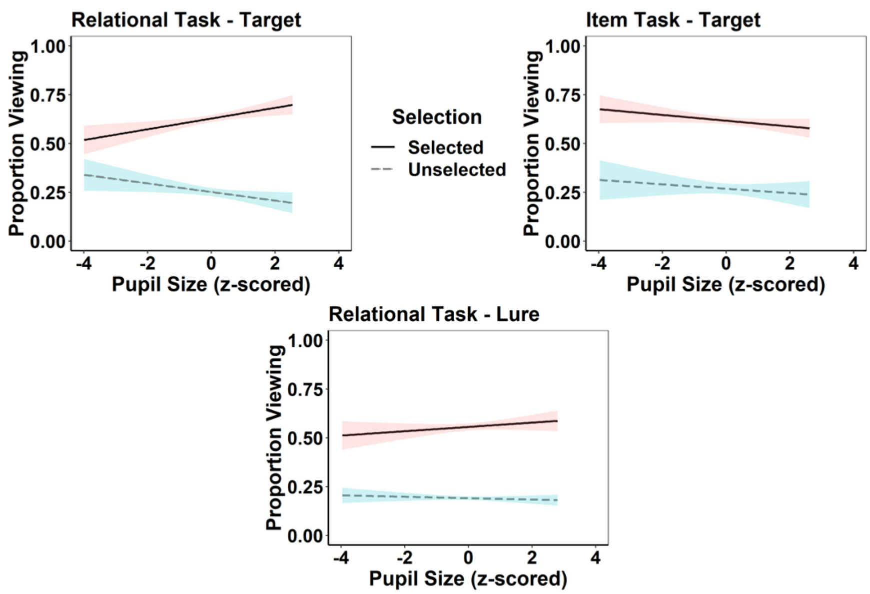
Number of Fixations to Visual Display Does Not Affect Pupil Size Between Tasks

Pupil sizes varied across the time course for both item and relational tasks, as shown in Figure 9. Given the differences in orienting instructions and scene manipulation between the item and relational task that presumably affected the manner in which participants processed the background scene during face–scene learning, we performed a follow-up analysis of whether the number of fixations on a given trial affected pupil size. To assess whether the number of fixations impacted pupil size on a trial-by-trial basis, a mixed effects regression model was fit to pupil size on each learning trial, using (number of) *fixations* and *task* (item vs. relational) as fixed effects and a random intercept for *participants*. The results are reported in Table 9. There was no effect of fixation, indicating that the number of fixations on each trial did not affect pupil size. There was no effect of task, indicating that the pupil size did not differ between the item task ($M = 0.03$, $SD = 0.96$) and the relational task ($M = -0.01$, $SD = -1.01$). The Fixation × Task interaction was also not significant, indicating that the relationship between number of fixations and pupil size did not differ between the item and relational task. This is especially important to rule out, given that there was a disproportionately greater amount of fixations in the relational task ($M = 3.62$, $SD = 2.36$) than the item task ($M = 1.98$, $SD = 1.58$). Therefore, despite the number of fixations differing between the item and relational task, this had no impact on the pupil size on each trial.

[6] Analysis was conducted identifying the time window in which viewing to targets peaked and stabilized, details of which are included in the online supplemental materials section.

Figure 7

Proportion of Viewing to Target Faces in the Relational and Item Task, and Lure Faces in the Relational Task, in the 1 s Time Bin Prior to Response, Across Standardized Pupil Size Differences During Encoding, in Experiment 3



Note. Greater values for pupil indicate greater pupil dilations at encoding. The units of pupil dilation, shown on the *x* axis, reflect *z*-scored values of the raw pupil size area. Error bands reflect 95% confidence intervals. See the online article for the color version of the figure.

Discussion

The purpose of Experiment 3 was to replicate the critical findings in Experiment 1, namely, that pupil size fluctuations during learning predicted the magnitude of viewing to targets as a function of whether they would be selected. In Experiment 3, where statistical power to detect the effects observed in Experiment 1 was increased, pupil size predicted viewing to targets in the relational task regardless of whether they would go on to be selected, albeit in opposite directions. Greater pupil size during encoding predicted a greater magnitude of viewing to selected targets and it also predicted reduced magnitude of viewing to unselected targets. Given that selecting a target constitutes a correct response whereas not selecting the target constitutes an incorrect response, the relationship between pupil size during encoding and viewing test items during retrieval was dependent on whether subsequent recognition was successful. Thus, we conducted a follow-up analysis of pupil size predicting subsequent accuracy while taking into account the manner in which test items were viewed. This additional analysis showed that pupil size predicted accuracy depending on the magnitude of viewing test items prior to their selection. Greater pupil size during encoding predicted subsequent hits when viewing targets was high, it predicted subsequent misses when viewing targets was low, and it was invariant across accuracy when viewing was moderate. In all cases, these effects were observed in the relational, but not the item, task. Importantly, in the relational task, the relationship between pupil and viewing was not observed for Lures, indicating that this relationship was not observed for merely all selected faces in the relational task. Thus, these effects were driven by memory for target faces and was not merely a selection effect.

Additionally, the overall difference in pupil size between the item and relational tasks was controlled for by including a luminance mask that helped to stabilize the size of pupil-dilation during presentation of the face–scene pairs. Importantly, after controlling for potential confounding issues we observed previously, the results obtained from Experiment 3 followed the same pattern as those of Experiment 1. Thus, our results were unlikely because of luminance confounds, differences in orienting instructions, or differences in the number of fixations between tasks, and are more likely reflecting differences in memory encoding processes between the tasks that influence the pupil dilations on each trial. Lastly, we found that the interaction of pupil dilation at encoding and viewing behavior at test was predictive of future recognition accuracy, but only in the relational task, not the item task.

Table 7

Outcomes of Mixed Effects Model Regression Analysis of Viewing to the Lure Faces in the Relational Task, Using Pupil and Selection (Whether the Test Item in the Analysis Would Go on to Be Subsequently Selected) as Fixed Effects and a Random Intercept for Participants (Experiment 3)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Pupil	<0.01	0.01	0.78	.436
Selection	0.37	0.01	35.48	<.001**
Pupil × Selection	0.02	0.01	1.65	.099

** $p < .001$.

Table 8

Outcomes of Mixed Effects Logistic Regression Analysis of Accuracy, Using Pupil, Task (Item vs. Relational), and Viewing as Fixed Effects and a Random Intercept for Participants (Experiment 3)

Fixed effect	Estimate	*SE*	*z* Value	*p* Value
Pupil	−0.08	0.06	1.39	.165
Task	0.85	0.10	8.28	<.001**
Viewing	2.47	0.10	26.37	<.001**
Pupil × Task	0.26	0.10	2.53	.012*
Pupil × Viewing	0.19	0.10	2.13	.034*
Task × Viewing	−0.44	0.19	2.38	.017*
Pupil × Task × Viewing	−0.48	0.18	2.59	.010*
Follow-up analyses to three-way interaction, split by task variable				
Item task				
Pupil	0.03	0.08	0.40	.688
Viewing	2.24	0.13	16.60	<.001**
Pupil × Viewing	−0.04	0.14	0.33	.744
Relational task				
Pupil	−0.21	0.08	2.76	.006*
Viewing	2.73	0.13	20. 36	<.001**
Pupil × Viewing	0.46	0.13	3.64	<.001**
Follow-up analysis to two-way interaction in relational task				
Viewing +1 *SD*	0.21	0.09	2.40	.020*
Viewing +0 *SD*	0.01	0.06	0.19	.850
Viewing −1 *SD*	−0.19	0.07	2.60	.010*

* $p < .05$. ** $p < .001$.

General Discussion

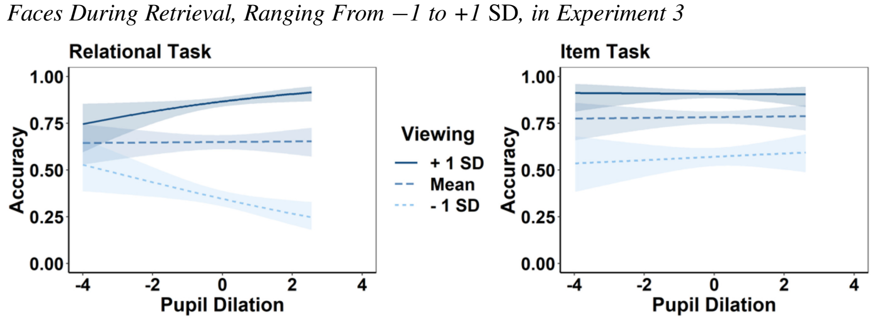
The critical question behind our investigation was whether the magnitude of viewing devoted to learned items at test was predicted by fluctuations in pupil size at the time of learning, and whether this differed between item and relational tasks. Across two studies we observed pupil size during encoding predicted the magnitude of viewing targets in the relational task, but not the item task. In Experiment 1, greater pupil size during encoding was related to greater viewing of subsequently selected targets in the relational task only. These effects were both replicated in Experiment 3 and, with increased statistical power, also showed the opposite relationship between pupil size and the magnitude of viewing to subsequently missed targets; namely, greater pupil size during encoding predicted reduced magnitude of viewing targets that would not go on to be selected. In both experiments, the relationship between pupil size at encoding and viewing at retrieval was not observed for lures in the relational task, regardless of whether they would be subsequently selected. Therefore, this pupil-viewing effect in the relational task did not generalize to any selected face, and instead was only the case for faces that were previously studied with the background scene. In addition, with increased statistical power in Experiment 3, pupil size during encoding was related to subsequent accuracy in the relational task only, dependent on the magnitude of viewing target items immediately prior to making a selection. When viewing to targets was high, pupil size predicted hits, whereas when viewing to targets was low, pupil size predicted misses. Intermediate levels of viewing target was not predictive of either hits or misses. Thus, the manner in which test items were viewed immediately prior to their selection was critical to the relation between pupil size and subsequent memory accuracy. In both sets of analyses, pupil size changes reflecting relational binding during learning were reflected in the magnitude of viewing target test items in the relational task, which in turn predicted subsequent accuracy.

Several follow-up analyses showed that the number of fixations during learning and the orienting instructions associated with each task had a negligible impact on pupil size on a trial-by-trial basis, ruling out the possibility that these observed relationships was because of differences in how the encoding display was viewed between the two tasks. Critically, the same effects observed in Experiment 1 when overall average pupil size differed between the two tasks were replicated in Experiment 3 when the overall average pupil size was similar between the two tasks. Therefore, differences in pupil size between the two tasks is insufficient in explaining the observed relationship between pupil size and viewing. Overall, our results suggest that pupillary fluctuations at encoding reflect the strength of having bound items to their studied contexts, and thus potentially serve as a marker for the strength of relational binding.

Our measure of relational memory in this study consisted of assessing the binding of faces with unique background scenes in memory and subsequently retrieving the specific association,

Figure 8

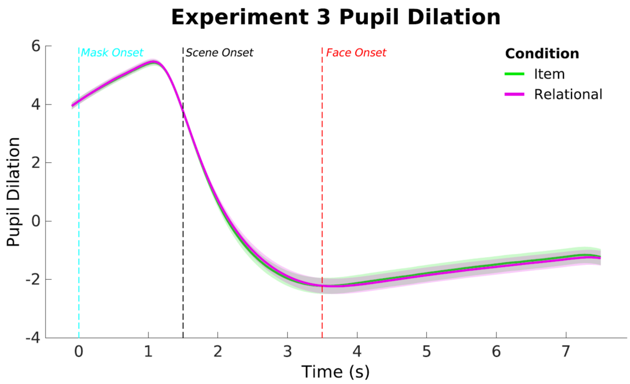
Accuracy in the Relational and Item Task Across Standardized Pupil Size Differences to Target Faces During Encoding as a Function of the Magnitude of Viewing to Target Faces During Retrieval, Ranging From −1 to +1 SD*, in Experiment 3*



Note. Greater values for pupil indicate greater pupil dilations at encoding. The units of pupil dilation, shown on the *x* axis, reflect *z*-scored values of the raw pupil size area. Error bands reflect 95% confidence intervals. See the online article for the color version of the figure.

Figure 9

Pupil Size Fluctuations Across the Full 6 s Time Course of Experiment 3



Note. An unobstructed scene was presented for 2 s followed by a face superimposed on that scene for 4 s. Greater values of pupil dilation indicate larger pupil sizes. The units of pupil dilation, shown on the *y*-axis, reflect *z*-scored values of the raw pupil size area. Error bands reflect standard error around the mean. See the online article for the color version of the figure.

whereas item memory consisted of encoding individual faces in the presence of a background scene that was inconsequential to later recognition of that face. The critical difference between the two tasks was that the item task simply relied on memory for individual faces, whereas the relational task required distinguishing between learned faces and background scenes. Pupil size fluctuations during encoding predicting viewing to targets in the relational task (but not the item task) likely indexed the engagement of memory processes meant to bind faces to the background scenes with which they were learned. Previous work with the relational task demonstrated that the strength of memory for selected targets was reflected in gradations in the magnitude of viewing behavior to those selected targets (Whitlock et al., 2020). Thus, the magnitude of viewing selected targets during retrieval in the relational task reflects the strength of associative binding of faces to their studied scenes. Results from the current study suggest that this associative binding is initially reflected in fluctuations in pupil size during encoding. Importantly, pupil sizes at encoding were not predictive of preferential viewing at test in the item task, and pupil sizes during encoding

Table 9

Outcomes of Mixed Effects Model Regression Analysis of Number of Fixations Relating to Pupil Size in Both the Item and Relational Tasks, Using (Number of) Fixations and Task (Item vs. Relational) as Fixed Effects and a Random Intercept for Participants (Experiment 3)

Fixed effect	Estimate	*SE*	*t* Value	*p* Value
Fixation	0.02	0.01	1. 70	.090
Task	−0.05	0.05	0.89	.373
Fixation × Task	0.03	0.02	1.54	.123
Effect of task on # of fixations				
Task	−1.65	0.06	27.13	<.001**

** $p < .001$.

did not predict the magnitude of viewing to lure faces in the relational task, regardless of whether they were selected, suggesting that pupil dilation predicts viewing behavior for faces at test only when they are presented with their originally encoded context. These results were corroborated with additional analyses predicting recognition memory accuracy using variations in pupil dilation at encoding. Namely, pupil dilations at encoding were predictive of future accuracy only in the relational task, not the item task.

Few studies have utilized pupillometry to examine ongoing encoding processes, and these studies have reported mixed results. For instance, Võ et al. (2008) found greater pupil dilation for studied words compared to new words during recognition testing, but no difference in pupil dilation during encoding that predicted future memory of the words. In contrast, Papesh et al. (2012) reported greater pupil dilation at encoding to words that were later remembered with high confidence compared to lower confidence ratings. Other studies using scene stimuli have found greater pupil dilation at encoding for subsequently missed stimuli compared to those successfully recognized, but memory strength was not assessed (Naber et al., 2013; Wetzel et al., 2020). Importantly, the previous studies all implemented single-item recognition memory paradigms, where retrieval of associated details was not necessary. However, high confidence responses may reflect recollection of specific or associated details of an episode (Yonelinas, 2001), and recent work suggests that pupil old/new effects during recognition are more greatly influenced by strategic and intentional uses of memory than automatic memory strength (Brocher & Graf, 2017). Given our results, we suggest that greater pupil dilations during the encoding of a particular stimulus does not simply reflect a generic creation of a memory or cognitive effort, but a more specific and intentional associative binding process that influences future memory decisions and viewing behavior, and may only be predictive of future memory success when that process is engaged. This may explain the mixed results of previous studies, and future studies may incorporate more associative or relational demands in memory tasks to more clearly identify the process reflected in pupil dilations.

Our interpretation of dynamic pupil dilations and relational binding is supported by neurophysiological studies linking pupil size to locus coeruleus–norepinephrine (LC–NE) activity in the brain (Aston-Jones & Cohen, 2005; Joshi et al., 2016), the primary neural region involved in synthesis of norepinephrine and dissemination of this neurotransmitter to other areas of the brain. Changes in pupil diameter are closely tied to neural firing of LC cells in both primates (Joshi et al., 2016) and humans (Murphy et al., 2011). Importantly, the LC–NE system has direct projections to the hippocampus, particularly to the dentate gyrus (Haring & Davis, 1985; Patton & McNaughton, 1995; Samuels & Szabadi, 2008), and neuromodulatory release from LC neurons to the hippocampus plays a critical role in spatial and contextual learning and memory, and inhibition of LC activity during learning greatly impairs successful encoding (Kaufman et al., 2020; Kempadoo et al., 2016; Wagatsuma et al., 2018). Greater subsequent memory effects in the hippocampus have been observed for more complex stimuli, such as objects in scenes, compared to more simple images such as individual objects (Kim, 2011), the former of which is also associated with greater exploratory viewing during learning (see Voss et al., 2017). The hippocampus is critical for encoding and retrieval of relational memories (Konkel & Cohen, 2009; Monti et al., 2012), and preferential viewing and efficient scanpath behavior during relational memory retrieval has been linked to hippocampal activity (Hannula et al., 2007; Hannula & Ranganath, 2009; Kragel et al., 2021; Lucas et al., 2019). Considering that the orienting instructions between the two tasks did not reliably elicit differences in pupil dilation, we believe that the nature of encoding or attentional effort explanation of our pupil dilation results is inadequate. Rather, it is possible that pupil dilation during encoding observed in our study might have reflected LC activity to hippocampus to promote contextual encoding, which led to preferential viewing of targets during test. We also note, however, that changes in pupil dilation are not a direct measure of LC–NE activity and can reflect contributions of other neural systems as well (Megemont et al., 2022; Yang et al., 2021). Further work combining eye-tracking and neuroimaging methodology will be necessary to evaluate this hypothesis.

To summarize, pupil size fluctuations during encoding predicted the magnitude of viewing to targets in the relational but not the item task in two experiments. Additionally, pupil size fluctuations during encoding predicted subsequent recognition memory success in the relational task as a function of the magnitude of viewing targets immediately prior to behavioral response. These effects were unlikely to be caused by luminance differences, encoding orientation, or fixations during the study period. The relational task involved binding of faces to their studied context, whereas the item task involved simply learning individual faces, and therefore the current findings suggest that fluctuations in pupil size serve as a marker for associative binding processes relating faces to their studied context. Furthermore, the strength of this associative binding was reflected in the subsequent magnitude of viewing to targets at test, where gradations in magnitude of viewing to selected targets was previously shown to reflect memory strength of those targets (Whitlock et al., 2020). Therefore, pupil dilations during learning may not simply reflect the strength of encoding of individual items but rather specific binding of items to the contexts with which they were learned.

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010. https://doi.org/10.1037//0278-7393.25.4.997

Ariel, R., & Castel, A. D. (2014). Eyes wide open: Enhanced pupil dilation when selectively studying important information. *Experimental Brain Research*, *232*(1), 337–344. https://doi.org/10.1007/s00221-013-3744-5

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baym, C. L., Khan, N. A., Pence, A., Raine, L. B., Hillman, C. H., & Cohen, N. J. (2014). Aerobic fitness predicts relational memory but not item memory performance in healthy young adults. *Journal of Cognitive Neuroscience*, *26*(11), 2645–2652. https://doi.org/10.1162/jocn_a_00667

Brocher, A., & Graf, T. (2017). Decision-related factors in pupil old/new effects: Attention, response execution, and false memory. *Neuropsychoilogia*, *102*, 124–134. https://doi.org/10.1016/j.neuropsychologia.2017.06.011

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165–178. https://doi.org/10.1016/j.visres.2015.03.005

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. The MIT Press.

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, *184*, 119–129. https://doi.org/10.1016/j.cognition.2018.12.014

DeBruine, L., & Jones, B. (2017). *Face research lab London set*. figshare. Dataset. https://doi.org/10.6084/m9.figshare.5047666.v5

Ding, H., Whitlock, J., & Sahakyan, L. (2021). *Cross-race effect in item and relational memory: An eye tracking investigation* [Manuscript submitted for publication]. Department of Psychology, University of Illinois at Urbana-Champaign.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain: 35*. Oxford University Press.

Eldar, E., Niv, Y., & Cohen, J. D. (2016). Do you see the forest or the tree? Neural gain and breadth versus focus in perceptual processing. *Psychological Science*, *27*(12), 1632–1643. https://doi.org/10.1177/0956797616665578

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

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

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, *6*(4), 316–328. https://doi.org/10.1002/(SICI)1097-0193(1998)6:4<316::AID-HBM9>3.0.CO;2-6

Gross, M. P., & Dobbins, I. G. (2021). Pupil dilation during memory encoding reflects time pressure rather than depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(2), 264–281. https://doi.org/10.1037/xlm0000818

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*, Article 166. https://doi.org/10.3389/fnhum.2010.00166

Hannula, D. E., Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The eyes know: Eye movements are a veridical index of memory. *Psychological Science*, *23*(3), 278–287. https://doi.org/10.1177/0956797611429799

Hannula, D. E., & Ranganath, C. (2009). The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron*, *63*(5), 592–599. https://doi.org/10.1016/j.neuron.2009.08.025

Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience*, *19*(10), 1690–1705. https://doi.org/10.1162/jocn.2007.19.10.1690

Haring, J. H., & Davis, J. N. (1985). Differential distribution of locus coeruleus projections to the hippocampal formation: Anatomical and biochemical evidence. *Brain Research*, *325*(1–2), 366–369. https://doi.org/10.1016/0006-8993(85)90342-7

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*(1), 119–129. https://doi.org/10.1111/j.1469-8986.2007.00605.x

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106. https://doi.org/10.3758/BF03195300

Hershman, R., Henik, A., & Cohen, N. (2018). A novel blink detection method based on pupillometry noise. *Behavior Research Methods*, *50*(1), 107–114. https://doi.org/10.3758/s13428-017-1008-1

Hess, E. H. (1965). Attitude and pupil size. *Scientific American*, *212*(4), 46–54. https://doi.org/10.1038/scientificamerican0465-46

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hess, E. H., Seltzer, A. L., & Shlien, J. M. (1965). Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology*, *70*(3), 165–168. https://doi.org/10.1037/h0021978

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition*, *27*(4), 657–664. https://doi.org/10.3758/bf03211559

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*(3), 341–358. https://doi.org/10.1037/0033-295X.94.3.341

Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*(1), 236–242. https://doi.org/10.1177/0956797613498260

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

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kahneman, D., Beatty, J., & Pollack, I. (1967). Perceptual deficit during a mental task. *Science*, *157*(3785), 218–219. https://doi.org/10.1126/science.157.3785.218

Kamp, S. M., & Zimmer, H. D. (2015). Contributions of attention and elaboration to associative encoding in young and older adults. *Neuropsychologia*, *75*, 252–264. https://doi.org/10.1016/j.neuropsychologia.2015.06.026

Kaufman, A. M., Geiller, T., & Losonczy, A. (2020). A role for the locus coeruleus in hippocampal CA1 place cell reorganization during spatial reward learning. *Neuron*, *105*(6), 1018–1026.e4. https://doi.org/10.1016/j.neuron.2019.12.029

Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., & Kandel, E. R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences*, *113*(51), 14835–14840. https://doi.org/10.1073/pnas.1616515114

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *Neuroimage*, *54*(3), 2446–2461. https://doi.org/10.1016/j.neuroimage.2010.09.045

Konkel, A., & Cohen, N. J. (2009). Relational memory and the hippocampus: Representations and methods. *Frontiers in Neuroscience*, *3*(2), 166–174. https://doi.org/10.3389/neuro.01.023.2009

Kragel, J. E., Schuele, S., vanHaerents, S., Rosenow, J. M., & Voss, J. L. (2021). Rapid coordination of effective learning by the human hippocampus. *Science Advances*, *7*(25), Article eabf7144. https://doi.org/10.1126/sciadv.abf7144

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, *3*(4), 525–551. https://doi.org/10.1016/0010-0285(72)90021-7

Lucas, H. D., Duff, M. C., & Cohen, N. J. (2019). The hippocampus promotes effective saccadic information gathering in humans. *Journal of Cognitive Neuroscience*, *31*(2), 186–201. https://doi.org/10.1162/jocn_a_01336

Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, *11*, Article e70510. https://doi.org/10.7554/eLife.70510

Mill, R. D., O'Connor, A. R., & Dobbins, I. G. (2016). Pupil dilation during recognition memory: Isolating unexpected recognition from judgment uncertainty. *Cognition*, *154*, 81–94. https://doi.org/10.1016/j.cognition.2016.05.018

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

Miller, A. L., & Unsworth, N. (2021). Attending to encode: The role of consistency and intensity of attention in learning ability. *Journal of Memory and Language*, *121*, Article 104276. https://doi.org/10.1016/j.jml.2021.104276

Molitor, J. J., Ko, P. C., Hussey, E. P., & Ally, B. A. (2014). Memory-related eye movements challenge behavioral measures of pattern completion and pattern separation. *Hippocampus*, *24*(6), 666–672. https://doi.org/10.1002/hipo.22256

Monti, J. M., Hillman, C. H., & Cohen, N. J. (2012). Aerobic fitness enhances relational memory in preadolescent children: The FITKids randomized control trial. *Hippocampus*, *22*(9), 1876–1882. https://doi.org/10.1002/hipo.22023

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Naber, M., Frässle, S., Rutishauser, U., & Einhauser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), Article 11. https://doi.org/10.1167/13.2.11

Nickel, A., Henke, K., & Hannula, D. (2015). Relational memory is evident in eye movement behavior despite the use of subliminal testing methods. *PLoS ONE*, *10*(10), Article e0141677. https://doi.org/10.1371/journal.pone.0141677

Norton, D., & Stark, I. (1971). Eye movements and visual perception. *Scientific American*, *224*(6), 34–43.

Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition*, *22*(7), 975–995. https://doi.org/10.1080/13506285.2014.941433

Olsen, R. K., Sebanayagam, V., Lee, Y., Moscovitch, M., Grady, C. L., Rosenbaum, R. S., & Ryan, J. D. (2016). The relationship between eye movements and subsequent recognition: Evidence from individual differences and amnesia. *Cortex*, *85*, 182–193. https://doi.org/10.1016/j.cortex.2016.10.007

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 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

Patton, P. E., & McNaughton, B. (1995). Connection matrix of the hippocampal formation: I. The dentate gyrus. *Hippocampus*, *5*(4), 245–286. https://doi.org/10.1002/hipo.450050402

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 1–12. https://doi.org/10.1167/9.10.2

Richardson, D. C., & Spivey, M. K. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

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 organization. *Current Neuropharmacology*, *6*(3), 235–253. https://doi.org/10.2174/157015908785777229

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Tullis, J. G., Benjamin, A. S., & Liu, X. (2014). Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & Cognition*, *42*(6), 863–875. https://doi.org/10.3758/s13421-014-0409-y

Unsworth, N., & Miller, A. L. (2021a). Encoding dynamics in free recall: Examining attention allocation with pupillometry. *Memory & Cognition*, *49*(1), 90–111. https://doi.org/10.3758/s13421-020-01077-7

Unsworth, N., & Miller, A. L. (2021b). Individual differences in the intensity and consistency of attention. *Current Directions in Psychological Science*, *30*(5), 391–400. https://doi.org/10.1177/09637214211030266

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review*, *22*(3), 757–765. https://doi.org/10.3758/s13423-014-0747-6

van der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). Visual memory for fixated regions of natural images dissociates attraction and recognition. *Perception*, *38*(8), 1152–1171. https://doi.org/10.1068/p6142

van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PLoS ONE*, *7*(12), Article e51134. https://doi.org/10.1371/journal.pone.0051134

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*(1), 130–140. https://doi.org/10.1111/j.1469-8986.2008.00745.x

Voss, J. L., Bridge, D. J., Cohen, N. J., & Walker, J. A. (2017). A closer look at the hippocampus and memory. *Trends in Cognitive Sciences*, *21*(8), 577–588. https://doi.org/10.1016/j.tics.2017.05.008

Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, *115*(2), E310–E316. https://doi.org/10.1073/pnas.1714082115

Wetzel, N., Einhäuser, W., & Wildmann, A. (2020). Picture-evoked changes in pupil size predict learning success in children. *Journal of Experimental Child Psychology*, *192*, Article 104787. https://doi.org/10.1016/j.jecp.2019.104787

Whitlock, J., Hubbard, R., Ding, H., & Sahakyan, L. (2023). *Trial-level fluctuations in pupil dilation during encoding reflect strength of relational binding*. https://osf.io/psf85/

Whitlock, J., Lo, Y. P., Chiu, Y. C., & Sahakyan, L. (2020). Eye movement analyses of strong and weak memories and goal-driven forgetting. *Cognition*, *204*, Article 104391. https://doi.org/10.1016/j.cognition.2020.104391

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Wynn, J. S., Olsen, R. K., Binns, M. A., Buchsbaum, B. R., & Ryan, J. D. (2018). Fixation reinstatement supports visuospatial memory in older adults. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(7), 1119–1127. https://doi.org/10.1037/xhp0000522

Yang, H., Bari, B. A., Cohen, J. Y., & O'Connor, D. H. (2021). Locus coeruleus spiking differently correlates with S1 cortex activity and pupil diameter in a tactile detection task. *eLife*, *10*, Article e64327. https://doi.org/10.7554/eLife.64327

Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1363–1374. https://doi.org/10.1098/rstb.2001.0939

Received August 23, 2021
Revision received June 28, 2023
Accepted June 29, 2023 ■