Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-022-02081-5

BRIEF REPORT

Pupillary correlates of individual differences in long-term memory

Matthew K. Robison[1] · Jamie M. Trost[2] · Daniel Schor[2] · Bradley S. Gibson[2] · M. Karl Healey[3]

Accepted: 4 March 2022

Abstract

The present study is the first to examine individual differences in long-term memory, arousal dysregulation, and intensity of attention within the same experiment. Participants (*N* = 106) completed 28 lists of an immediate free-recall task while their pupil diameter was recorded via an eye-tracker during the encoding period. Two main pupillary measures were extracted: intraindividual variability in pre-list pupil diameter and evoked pupillary responses during item encoding. Variability in pre-list pupil diameter served as a measure of arousal dysregulation, and evoked pupillary responses served as a measure of intensity of attention. Based on prior work, we hypothesized that there would be a positive association between intensity of attention and recall ability, and that there would be a negative association between arousal dysregulation and recall ability. Collectively these two measures accounted for 19% of interindividual variance in recall, with 5% attributable uniquely to intensity of attention and 12% attributable uniquely to arousal regulation. The findings demonstrate that there are sources of individual differences in long-term memory that can be revealed via pupillometry, notably the amount of effort deployed during item encoding and the degree to which people exhibit dysregulated arousal. Both findings are consistent with recent theorizing regarding the role of the locus coeruleus (LC)-norepinephrine (NE) system's role in goal-directed cognition. Specifically, the LC governs both moment-to-moment arousal and NE release to cortical regions subserving cognitive processing. Among people for whom this system operates most optimally, long-term memory retention is superior.

Keywords Free recall · Long-term memory · Pupillometry

Introduction

For over 100 years, cognitive psychologists have been studying why people differ in their ability to encode information into long-term memory (see Unsworth, 2019, for a recent review). Several candidate sources of variability have been identified. One is the degree to which individuals can perform a controlled search of memory for relevant information, including how well they organize memories according to the temporal order in which the information was encoded, how well they use semantic cues to find relevant information, how efficiently they search memory for target information, and how well they monitor the outputs of the memory search process (Healey et al., 2014; Miller & Unsworth, 2018; Spillers & Unsworth, 2011; Unsworth et al., 2013; Unsworth & Engle, 2007; Unsworth et al., 2011). Researchers have also leveraged pupillometry to understand individual differences in the outlay of effort toward encoding information and relative functioning of the locus coeruleus-norepinephrine (LC-NE) system (Madore et al., 2020; Miller et al., 2019). Here we leverage pupillometry to investigate these two additional potential sources of interindividual variation in long-term memory: intensity of attention and arousal (dys)regulation.

Kahneman and Beatty (1966) were the first to demonstrate that pupil diameter was sensitive to the outlay of effort toward encoding and retrieving information in memory. A host of subsequent studies have also found that the pupil dilates in response to the encoding of information, either for maintenance in working memory (Alnæs et al., 2014; Aminihajibashi et al., 2020; Heitz et al., 2008; Kursawe & Zimmer, 2015; Meghanathan et al., 2015; Robison & Unsworth, 2019; Siegle et al., 2003; Unsworth & Robison, 2015, 2018) or later retrieval from long-term memory (Ariel & Castel, 2014; Gross & Dobbins, 2021; Kahneman & Peavler,

✉ Matthew K. Robison
matthew.robison@uta.edu

1 University of Texas at Arlington, 313 Life Science Building, 501 Nedderman Drive, Box 19528, Arlington, TX 76019, USA

2 University of Notre Dame, Notre Dame, IN, USA

3 Michigan State University, East Lansing, MI, USA

Published online: 30 March 2022

1969; Miller & Unsworth, 2020, 2021; Miller et al., 2019; Papesh et al., 2012; Peavler, 1974; Unsworth & Miller, 2021). At the level of individual differences, Miller et al. (2019) found a positive correlation between evoked pupillary responses during encoding and performance on delayed free recall. Similarly, Miller and Unsworth (2020) showed a positive correlation between pupillary responses at encoding and performance on a paired-associates task. This has led Miller and Unsworth to propose that intensity of attention is an important individual difference that can partially account for why people differ in long-term memory abilities.

Researchers have also been leveraging variation in arousal (both within and across people) to understand individual differences in cognition. Specifically, Unsworth and Robison recently proposed that arousal regulation may serve as a crucial individual difference variable underlying working memory capacity and attention control – two abilities that are also important correlates of long-term memory abilities (Kane & Engle, 2000; Rosen & Engle, 1997; Unsworth & Spillers, 2010). At least to a certain extent, arousal is driven by activity in the LC-NE system. The LC is a small brainstem nucleus that releases most of the NE into cortex (Aston-Jones & Cohen, 2005; Berridge & Waterhouse, 2003). As such, it is largely responsible for moment-to-moment arousal levels. Unsworth and Robison (2017a) proposed that individuals may differ in the stability of these moment-to-moment arousal levels, with greater variability reflecting less regulation, and that the extent of this arousal regulation may have consequences for individual differences in cognitive ability (see also, Tsukahara et al., 2016).

The LC is difficult to study in humans because of its small size and location. But recently, it has become evident that pupil diameter can be used as a proxy for LC activity (Aston-Jones & Cohen, 2005; Joshi & Gold, 2020; Joshi et al., 2016; Varazzani et al., 2015). Consistent with this notion, several recent studies have measured arousal regulation by measuring variability in pupil diameter and found relations among arousal regulation, long-term memory, attention control, sustained attention, working memory capacity, self-reported instances of mind-wandering and distraction, and self-reported media multitasking (Aminihajibashi et al., 2020; Aminihajibashi et al., 2019; Madore et al., 2020; Robison & Brewer, 2020, in press; Robison & Unsworth, 2019). Therefore, it appears that arousal regulation is a task- and domain-general individual difference that can impact a host of cognitive performance measures.

The present study

Here we test the hypothesis that arousal regulation is an important individual difference variable for long-term memory. This hypothesis has been difficult to examine in previous studies because the memory tasks did not include a sufficient number of trials in which to measure variability (see e.g., Miller et al., 2019; Miller & Unsworth, 2020, 2021). The one exception was a study conducted by Madore et al. (2020) in which participants performed 252 trials of recognition memory. The main results showed a negative correlation between arousal dysregulation (trial-to-trial variability in pupil diameter) and memory performance (*d'*). To our knowledge, Madore et al.'s study is the first to examine the association between arousal regulation and long-term memory. The present design allowed us to examine the relative contributions of both intensity of attention and arousal regulation to individual differences in long-term memory. This study will be the first to examine these two aspects within the same sample.

Our goal here was to extend the task given by Miller et al. (2019) so that both arousal dysregulation and intensity of attention could be reliably measured. Participants completed 28 lists of 12 words in an immediate free-recall task, presumably enough lists to enable us to observe fluctuations in arousal across the course of a 1-h session. Based on the LC-NE theory of individual differences, we predicted that dysregulation of arousal (measured via variability in pre-list pupil diameter), would correlate with *lower* average recall. Additionally, based on the work of Miller et al. (2019, Miller & Unsworth, 2020; Unsworth & Miller, in press), we predicted that greater evoked pupillary responses at encoding would correlate with *higher* average recall. Various factors can elicit smaller or larger evoked pupillary responses when people encode information. For example, unexpected memoranda are accompanied by greater evoked pupillary responses at encoding (Frank & Kafkas, 2021; Kafkas & Montaldi, 2018). There are also perceptual influences on pupillary responses. For example, words that carry "bright" meanings produce relative pupillary constrictions and words that carry "dark" meanings produce relative pupillary dilations. Of course, any perceptual differences can affect pupillary responses as well (e.g., physically brighter words will produce pupillary constriction relative to physically darker words; Mathôt et al., 2017). In the present study, we are most interested in endogenously produced effortful attention brought to bear by the observer. A participant can exert more or less attention to encoding any given word. Our supposition is that individual differences in pupillary responses in the present study are largely due to these endogenous factors – some participants exert more attention at encoding than others. Presumably, all other factors that affect pupillary responses (e.g., physical luminance differences, brightness connotations, novelty) would be relatively even across participants.

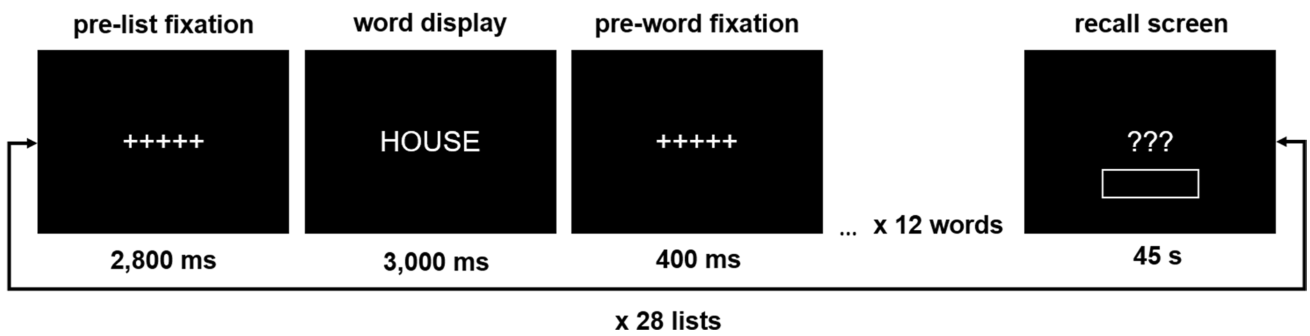

Fig. 1 Diagram of immediate free-recall task. Participants completed 28 lists of 12 words each. Each list started with a 2,800-ms fixation screen. Each word was presented for 3,000 ms with a 400-ms fixation screen presented between each word. Immediately after the presentation of the 12th word, participants were prompted to type their responses into a text box on the screen. Participants received 45 s to recall the list

Method

Participants and procedure

A sample of 119 participants from the human subject pool at the University of Notre Dame completed the study in exchange for partial course credit. To achieve 80% power, a minimum sample size of 90 was required for $\alpha = 0.05$ to detect a correlation of 0.30. We administered the task to as many participants as possible during a single academic semester, using the end of the semester as the discontinuation point for data collection. Prior to study commencement, participants provided informed consent. Then, they were seated in front of a computer with an eye-tracker mounted to the bottom of the monitor. Participants sat about 60 cm from the screen, freely viewing without a chinrest. The lights in the experimental room were dimmed to a constant setting for all participants. Thirteen participants were excluded from the analysis because there were technical issues with the computer/eye-tracker during calibration or the experiment. Participants completed a 3-min pre-experimental baseline measure during which they stared at a white fixation cross against a black background. Then they completed the immediate free-recall (IFR) task.

Immediate free-recall task

Participants completed 28 lists of an immediate free recall (IFR) task. Each list began with a 2,800-ms fixation screen. Then, participants saw a list of 12 words. Each word was presented for 3,000 ms separated by a 400-ms fixation screen. After the final word, a recall screen appeared (Fig. 1). Participants had 45 s to recall as many words from that list as possible by typing them into the computer. They were allowed to recall the words in any order they chose. Lists were presented in seven-list blocks. At the end of each block, participants were allowed to take a break and self-initiate the next block. The task took an average of 50 min to complete ($SD = 2$ min). The recall data were scored by marking responses as correct if a recalled word was indeed presented on the immediately preceding encoding list. Repetitions (correct responses that were provided more than once), previous-list intrusions (words recalled that were presented on prior lists), and extra-list intrusions (responses that were not on any list) were not marked as correct responses. For each list, we computed a recall proportion (correctly recalled items/12) and then averaged this proportion across the 28 lists.[1] This value was used in the analyses as the recall score for each participant.

Pupillometry

A Tobii eyetracker mounted to the computer monitor continuously recorded pupil diameter and gaze position data for both eyes at a sampling rate of 120 Hz. Participants' eyes were calibrated using a 9-point calibration screen at the beginning of the experiment. The pupil from the right eye was used (left and right eye measurements correlated at $r = 0.93$). Missing data due to blinks and off-screen fixations were excluded from the analysis (see *Results* for an analysis of missing data). We computed two main dependent variables: evoked pupillary responses (intensity of attention) and list-to-list variability in pupil diameter (arousal dysregulation). There are multiple ways to compute evoked pupillary responses in memory tasks. For example, some

[1] Previous-list intrusions and extra-list intrusions were rare (on average, 2% of all responses were extra-list intrusions, and 0.4% of all responses were previous-list intrusions). These responses were excluded from recall proportions (e.g., if a participant correctly recalled 6 out of 12 words, plus a word that was not on the list, their recall score for that list was 0.50).

Table 1 Descriptive statistics

Measure	Mean	SD	Skew	Kurtosis	Reliability
Recall accuracy	0.62	0.15	-0.10	-0.81	0.97
Mean pre-list pupil diameter	5.21	0.73	-0.20	0.57	0.99
CoV pre-list pupil diameter	0.08	0.03	0.80	0.28	0.82
Mean word-evoked pupil diameter (z)	0.18	0.19	0.58	0.76	0.71

Note. N = 106, SD = standard deviation, CoV = coefficient of variation. Reliabilities for pre-experimental measures were computed with split halves (first 90 s, second 90 s). Reliabilities for recall, prelist pupil measures, and word-evoked pupillary responses were computed with odd-list/even-list split halves. Reliability was then computed using the Spearman-Brown split-half formula

studies have subtracted samples from a pre-presentation fixation screen and reported the changes in millimeters (Ariel & Castel, 2014; Miller & Unsworth, 2020, 2021; Miller et al., 2019; Papesh et al., 2012; Unsworth & Miller, 2021). Other studies have standardized pupil diameter within a list, then examined relative pupil diameters after the onset of each word compared to the fixation screen (Kucewicz et al., 2018; Wainstein et al., 2017). We examined evoked responses with both methods, and the waveforms were nearly identical in shape and timecourse. However, the standardizing method reduced noise in the measurement, probably due to a reduction in intra- and interindividual variability. The same pattern of results was observed using both methods, and the two measures correlated highly ($r = 0.87$) at the participant level. But the standardization method had a stronger correlation with recall at the between-participant level. This standardized measure is reported in the *Results*, but analyses using both methods are reported in the Online Supplemental Materials. To measure arousal regulation, we also computed pre-list pupil diameter by averaging pupil diameter over the 2,800-ms fixation screen preceding each of the 28 lists. This measure was subsequently used to compute intraindividual variability in prelist pupil diameter. For each pre-list measurement, all available pupil diameter values for the 2,800-ms window were averaged. On some trials, there were no valid measurements, and these trials was excluded from the analysis. We created a variable for missingness as a sum of lists for which there was no available pre-list pupil data.

Data analysis

We used R (R Core Team, 2017) for all our analyses. To aggregate, transform, and plot data, we used the *tidyverse* (Wickham, 2017), *data.table* (Dowle & Srinivasan, 2018), and *cowplot* (Wilke, 2019) packages. We used the *lmerTest* (Kuznetsova et al., 2017) package to specify and estimate significance for parameter estimates in mixed-effect models, and we used the *EMAtools* package (Kleiman, 2017) to estimate effect sizes for mixed-effect models. For all dependent variables, we screened outlying data points by excluding anything outside 3 standard deviations of the mean. The article was written using the *papaja* (Aust & Barth, 2018) package. The data and analysis script are available publicly on the Open Science Framework at the following URL: https://osf.io/275em/

Results

Descriptive statistics for all participant-level measures are listed in Table 1. The first set of analyses focused on the eye-tracking measures. We were specifically interested in what measures correlated with recall performance at the within- and between-participant level. The first set of analyses examined pupillary dynamics within the context of the IFR task. We extracted two measures: variability (CoV) in prelist pupil diameter and mean word-evoked pupillary response. The measures were designed to capture fluctuations in arousal across the course of the task and the intensity of attention at encoding, respectively. Average prelist pupil diameter is plotted as a function of list in Fig. 2. As can be seen, prelist pupil diameter systematically declined across lists ($b = -0.01$, $SE = 0.001$, $p = < 0.001$, $d = -0.64$). This is consistent with prior work examining pupil diameter as a function of time-on-task (Hopstaken et al., 2015a, 2015b; Hopstaken et al., 2016; Massar et al., 2016; Unsworth & Robison, 2016). Figure 4 also reveals that measurements immediately following breaks (lists 1, 8, 15, and 22) are much lower than measurements preceding other lists. However, excluding these measurements led to virtually identical measurements of mean and variability of prelist pupil diameter.[2] For each participant, we computed the CoV of prelist pupil diameter across the 28 measurements. These measures were used for the analyses of individual differences.

The evoked pupillary responses are plotted in Fig. 3. As can be seen in Fig. 3A, pupil diameter quickly

[2] Mean prelist pupil diameter including and excluding lists following breaks correlated at 0.998, variability in prelist pupil diameter correlated at 0.975.

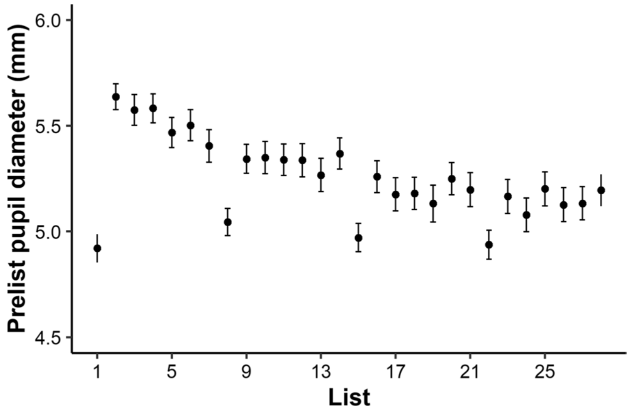

Fig. 2 Prelist pupil diameter by list, averaged across all participants. Pupil diameter immediately following breaks (lists 1, 8, 15, and 22) were lower than other lists. Error bars represent ± 1 standard error of the mean

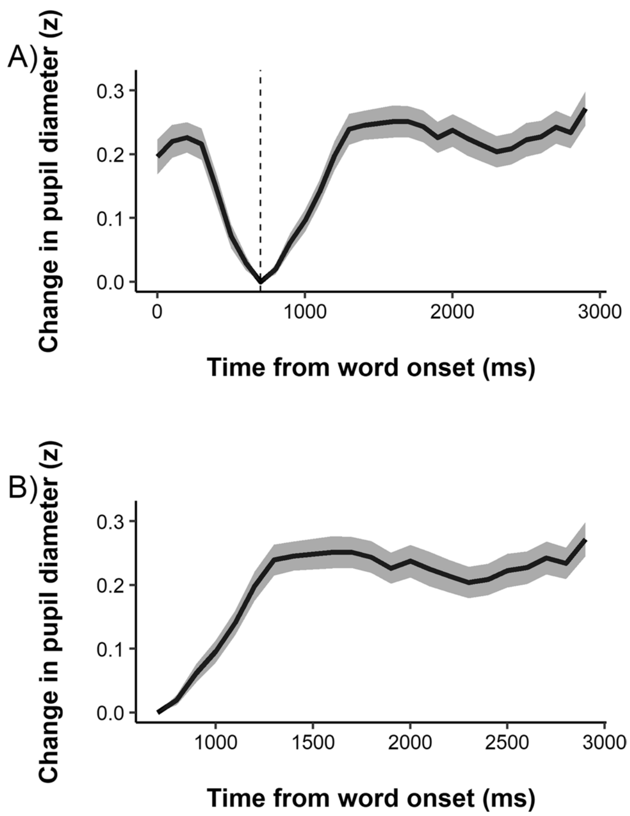

Fig. 3 (**A**) Evoked pupillary response baseline-corrected to the pre-word fixation interval, and (**B**) evoked pupillary response baseline-corrected to the period 700-ms after the word appeared to account for the pupillary light reflex. The average dilation over the window from 700 ms to 3,000 ms after word onset was used for analyses of individual differences. Shaded error bars represent ± 1 standard error of the mean

constricted following the onset of a word, presumably reflecting the pupillary light reflex. Then, starting at 700 ms after word-onset, the pupil begins to dilate and sustain a dilation throughout the remainder of the encoding window. Therefore, we re-baselined the evoked pupillary responses to the 700-ms timepoint (Fig. 3B; see Miller et al., 2019, for a similar method). Then, we averaged the change in pupil diameter on an item-by-item basis over the window from 700 ms to 3,000 ms after word onset, then averaged the item-level data for each participant within a list and across lists. The participant-level value was used for the analyses of individual differences. Item-level and list-level data were used to examine subsequent memory effects.

Based on prior work, we predicted that variability in arousal would negatively correlate with recall performance (Madore et al., 2020; Robison & Brewer, 2020; Robison & Unsworth, 2019; Unsworth & Robison, 2015, 2017b), and that intensity of attention would positively correlate with recall performance (Miller & Unsworth, 2020, 2021; Miller et al., 2019). Both hypotheses were supported by the data. Specifically, arousal dysregulation negatively correlated with recall accuracy ($r = -0.37$, $p < 0.001$; Fig. 4A, Table 2), and intensity of attention positively correlated with recall accuracy ($r = 0.26$, $p = 0.008$; Fig. 4B).

Next, to further investigate how arousal dysregulation and intensity of attention affected specific aspects of recall (primacy items, middle-list items, recency items), we submitted recall probability to a mixed model with a fixed, quadratic effect of serial position. The model revealed a significant quadratic effect of serial position ($b = 0.004$, $SE = 0.0001$, $p < 0.001$), typical of IFR responses. Words presented at the beginning and end of the list were recalled with greater likelihood than words in the middle of the list (primacy and recency effects, respectively). Then, we entered prelist CoV as a continuous fixed effect that was also allowed to interact with serial position. This model revealed a significant main effect of prelist CoV ($b = -0.08$, $SE = 0.01$, $p < 0.001$), and a significant prelist CoV × serial position interaction ($b = 0.004$, $SE = 0.0001$, $p < 0.001$). Arousal dysregulation had a larger effect on primary items than recency items (see Fig. 5A).

We repeated the above analysis using average TEPR as a continuous fixed effect, rather than prelist CoV. There was a significant main effect of TEPR on recall ($b = .04$, $SE = .01$, $p = .006$), but there was not a significant TEPR x serial position interaction ($b = 0.0004$, $SE = 0.0001$, $p = 0.41$). Intensity of attention thus had a relatively equal effect on recall of items at all serial positions (see Fig. 5B).

Finally, to examine whether these measures accounted for shared or independent sources of variance, we entered the two measures into a multiple regression predicting recall accuracy (see Table 3). Both arousal dysregulation *and* intensity of attention uniquely accounted for significant portions of variance in recall performance (see Table 2).

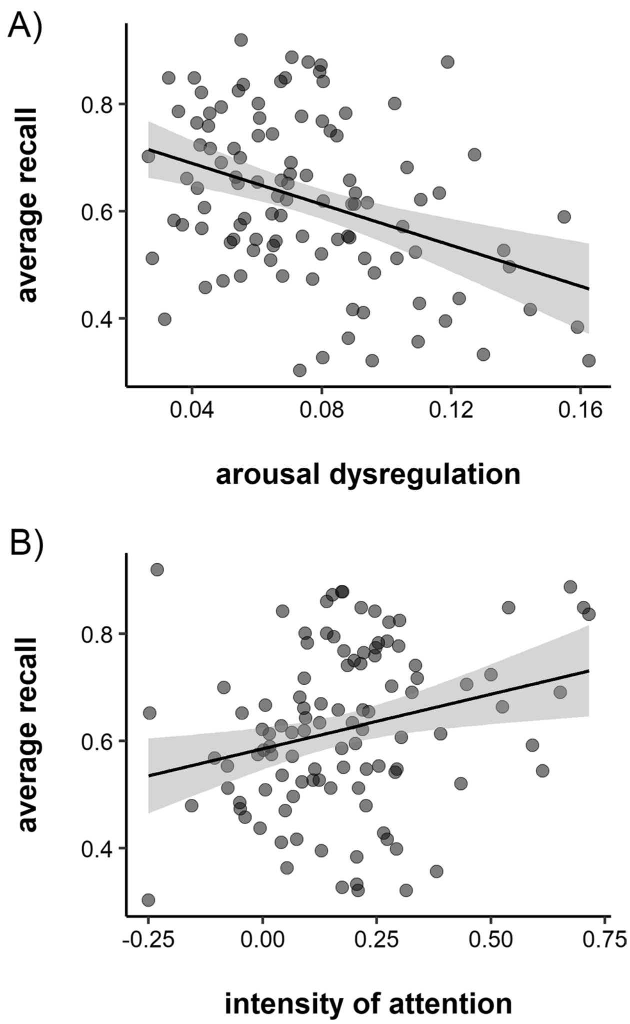

Fig. 4 (**A**) Scatterplot of the correlation between arousal dysregulation (variability in prelist pupil diameter) and average recall, and (**B**) scatterplot of the correlation between intensity of attention (average word evoked response) and average recall. The solid line represents the line of best fit through the points with the associated standard error in grey

Table 2 Correlations among recall and in-task pupillary measures

	1	2	3
1. Recall	–		
2. Arousal dysregulation	-0.37*	–	
3. Intensity of attention	0.26*	-0.10	–

Note. $N = 106$, $*p < 0.05$

Collectively, the two measures accounted for 19% of the total variance in recall. Thus, the data suggest that both dysregulation of arousal and intensity of attention partially account for individual differences in memory abilities. Importantly, intensity of attention and arousal dysregulation were not significantly correlated, suggesting these sources of variance are independent and manifest as distinct individual differences.[3]

One potential reason for list-to-list variability in pupil diameter may be missing data. That is, participants with more missing pupil data may end up showing larger values for prelist CoV. To examine this issue, we examined correlations between missing data, prelist CoV, TEPRs, and recall. There was indeed a positive correlation between missingness and prelist CoV ($r = 0.43$, $p < 0.001$), and a negative correlation between recall and missingness ($r = -0.45$, $p < 0.001$). Missingness did not significantly correlate with TEPR ($r = -0.09$, $p = 0.35$). When entered into a multiple regression predicting recall performance, all three independent variables (prelist CoV, average TEPR, and missingness) accounted for significant portions of variance in recall (TEPR: $\beta = 0.22$, $p = 0.02$; prelist CoV: $\beta = -0.21$, $p = 0.03$; missingness: $\beta = -0.42$, $p < 0.001$, $R^2 = 0.27$). Therefore, although there was certainly an association between missingness and arousal dysregulation, and between missingness and recall, this did not entirely account for the relation between recall and arousal dysregulation. The relation between missingness and recall is interesting, but it is unclear what could be driving this effect. Clearly, if participants are not looking at the screen, they will not encode the words, nor will the eye-tracker be able to collect data from their eyes. But data can be missing for several different reasons (e.g., eyes wandering off-screen, rubbing one's eyes, blinking, looking down at the keyboard, eye-tracker malfunction). So, it is difficult to know what precisely is driving this relation.

Subsequent memory effects

Some prior studies have found larger pupillary responses at both encoding and retrieval for items that are ultimately recalled versus forgotten (also called *subsequent memory effects*). For example, Kucewicz et al. (2018) observed larger pupillary responses to the encoding of subsequently recalled versus forgotten items in a delayed free-recall task. Similarly, Papesh et al. (2012) observed larger pupillary responses at encoding for words that were confidently recognized as studied versus other items in a recognition memory task. However, these effects are not always observed. For

[3] It is worth noting that in some prior studies (Robison & Unsworth, 2019; Unworth & Robison, 2017b), variability in evoked pupillary responses correlate with, and account for separable variance in, performance in working memory and attention tasks. That was not the case here, as intraindividual variability in evoked responses did not correlate with recall ($r = -0.09$, $p = 0.36$), and adding it to the regression model did add any additional attributable variance in recall.

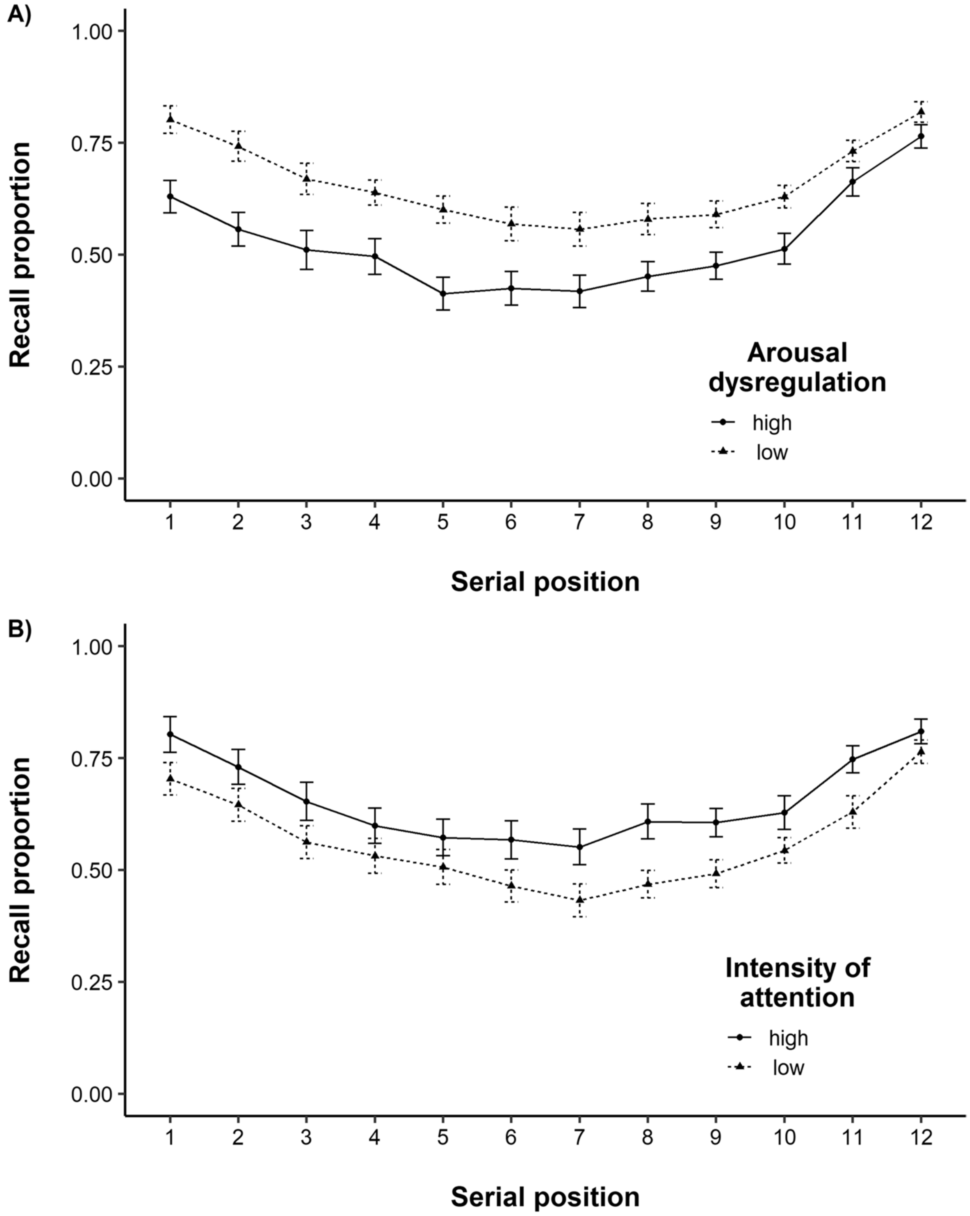

Fig. 5 Recall by serial position for participants with (**A**) high arousal dysregulation (highest quartile) and low arousal dysregulation (lowest quartile), and (**B**) high intensity of attention (highest quartile) and low intensity of attention (lowest quartile). Error bars represent ± 1 standard error. *Note:* Although upper and lower quartiles are plotted, arousal dysregulation and intensity of attention were treated as continuous variables in the analyses

Table 3 Regression on recall performance with in-task pupil measures

	β	SE	*t*	*p*
Arousal dysregulation	-0.35	0.09	- 3.95	< 0.001
Intensity of attention	0.22	0.09	2.45	0.02

Note. DV = recall accuracy. $R^2 = 0.19$. All variables are continuous and standardized. CoV = coefficient of variation

example, Unsworth and Miller (2021) did not find subsequent memory effects during a delayed free-recall task in any of their four experiments, and Gross and Dobbins (2021) did not observe subsequent memory effects in a recognition task. In fact, some studies have found reverse effects, with remembered information showing significantly smaller pupillary responses at encoding than forgotten information (Kafkas & Montaldi, 2011) or larger constriction to remembered images compared to forgotten images (Naber et al., 2013). Here, we examined both item-level and list-level

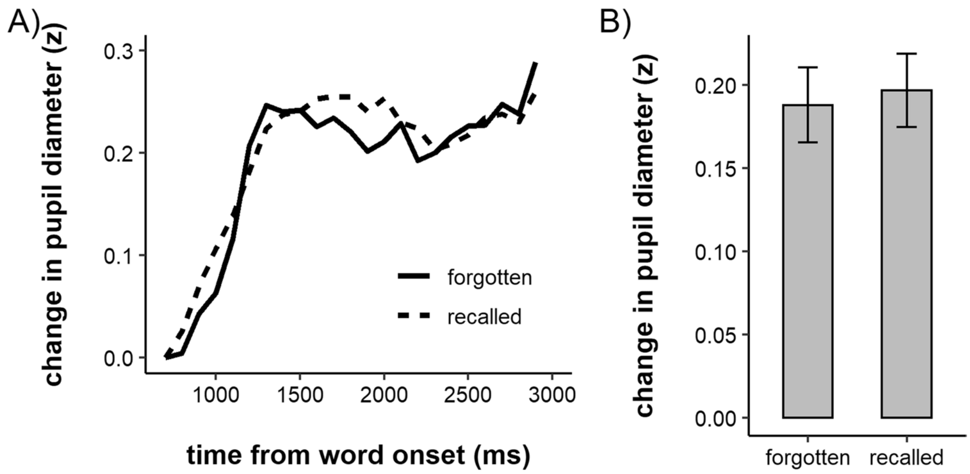

Fig. 6 (**A**) Evoked pupillary responses for recalled and forgotten words, and (**B**) average dilation for recalled and forgotten words. Error bars represent ± 1 standard error of the mean

subsequent memory effects. That is, are *words* accompanied by greater evoked pupillary responses at encoding ultimately remembered better? Also, are *lists* accompanied by greater pupillary responses at encoding remembered better? To do so, we specified a logistic regression with outcome (0 = forgotten, 1 = remembered) as the dependent variable, the evoked pupillary response for each trial as a fixed effect, and participant as a random effect. The average waveforms for recalled and forgotten words are plotted in Fig. 6A, and the average dilation for the waveform is plotted in Fig. 6B. As can be seen in the figure, the waveforms were quite similar, and the average dilation did not differ for remembered versus forgotten words ($b = 0.004$, $SE = 0.005$, $p = 0.38$). Likewise, the average evoked response did not differ for lists that were remembered better ($b = 0.01$, $SE = 0.01$, $p = 0.11$). Thus, although we observed evidence for participant-level effects, we did not observe evidence for item- or list-level effects of evoked pupillary responses on recall.

Discussion

The present study examined individual differences in long-term memory using pupillometry in an immediate free recall task. Based on prior studies finding correlations between recall performance and pupillary measures of intensity of attention (Miller & Unsworth, 2020, 2021; Miller et al., 2019), we hypothesized that larger average evoked pupillary responses at encoding would correlate with better recall. Also, based on prior work showing correlations between arousal dysregulation and cognitive ability measures like attention control, working memory capacity, and recognition memory (Aminihajibashi et al., 2019, 2020; Madore et al., 2020; Robison & Brewer, 2020; Robison & Unsworth, 2019; Unsworth & Robison, 2015, 2017b), we expected arousal dysregulation (i.e., more list-to-list variation in pupil diameter) to correlate with lower recall. As hypothesized, both arousal dysregulation ($r = -0.37$) and intensity of attention correlated with recall ($r = 0.26$). Further, in a multiple regression, both arousal dysregulation and intensity of attention accounted for significant portions of variance in recall performance, suggesting these are distinguishable individual differences.

These findings replicate and extend recent work using pupillometry to investigate potential reasons for individual differences in cognitive abilities. Although much prior work has shown correlations between working memory capacity, attention control, and arousal dysregulation, to our knowledge only one study has specifically addressed the relation between arousal dysregulation and long-term memory (Madore et al., 2020). However, Madore et al. examined arousal dysregulation during both encoding and retrieval of an incidental-encoding recognition memory paradigm. Arousal dysregulation during both encoding and retrieval correlated with lower memory performance. Here we examined arousal dysregulation during intentional encoding of to-be-remembered information. However, we did not design the task to be able to measure pupillary dynamics during retrieval (participants were allowed to look down at a keyboard to type their responses.) Therefore, we could not examine either arousal dysregulation or intensity of attention during retrieval in the present study. However, the combination of the present results and those from Madore et al. (2020) suggest that arousal dysregulation is a general characteristic that can exert its influence both at encoding *and* retrieval during both free recall and recognition memory.

Collectively the data are consistent with a framework recently outlined by Unsworth and Miller (in press). They argue that there are individual differences in the intensity with which people allocate their attention, which in the present study was measured by pupillary dilations during encoding, and the consistency with which they attend to a task from moment to moment, which we measured via prelist pupil variability. They argue that these are distinct individual differences, and our data are consistent with this argument. Overall, it appears that these two individual differences are important for a host of cognitive abilities including attention control (Unsworth & Robison, 2017a, 2017b; Unsworth, Miller, & Robison, 2020), working memory capacity (Robison & Brewer, 2020; Robison & Unsworth, 2019; Unsworth & Robison, 2015, 2017a, 2017b), and long-term memory (Madore et al., 2020; Miller & Unsworth, 2020, 2021). Further, consistency and intensity can both be measured covertly via pupillometry.

Although we observed a correlation between evoked pupillary responses and recall at the participant level, we did not observe item-level or list-level subsequent memory effects. The evidence for these effects is rather mixed. Papesh et al. (2012) found item-level effects in a recognition paradigm, and Kucewicz et al. (2018) found item-level effects in a delayed free-recall task. In working memory tasks, evoked pupillary responses can reveal the quantity of information held in memory, as well (Robison & Unsworth, 2019). But Unsworth and Miller (2021) did not find subsequent memory effects across four different delayed free-recall tasks that varied in presentation duration and list length. Several studies using recognition memory tests have also observed null effects (Gross & Dobbins, 2021), or patterns in the opposite direction (Kafkas & Montaldi, 2011; Naber et al., 2013). So, although one might expect to see better recall for words that are encoded more intensely, that was not the case here. Recently, Gross and Dobbins (2021) have argued that evoked pupillary dilations during item encoding might reflect time pressure induced by the limited exposure duration, rather than effort toward effectively encoding the words. It is worth noting that several studies that have observed significant subsequent memory effects (Kafkas & Montaldi, 2011; Naber et al., 2013; Papesh et al., 2012) used recognition memory paradigms where participants had to categorize items as old or new, whereas to our knowledge only one study has shown subsequent memory effects with delayed or immediate free recall (Kucewicz et al., 2018). Clearly, more research is needed on this phenomenon.

Future directions

In future work, several unanswered questions should be addressed. First, why is it the case that some people show greater intensity of attention at encoding? Is it because they are using more elaborative encoding strategies? Is it because they are simply applying more effort toward the task? Unfortunately, we did not collect any information regarding strategies or motivation from participants, so we could not answer those questions here. Previously, Miller and Unsworth (2020, 2021) showed that the effect of evoked pupillary responses on memory remained after controlling for other individual differences like encoding strategies and working memory capacity. So, although it is possible that something like motivation is contributing to the individual differences in evoked responses, it is unlikely that this covariation is solely due to encoding strategies. More work is needed on what drives individual differences in intensity of attention. Second, why is it the case that some people show relatively dysregulated arousal? Again, this could be a state-related source of variation, driven by something like motivation or fatigue. Or, it could be a stable, trait-level individual difference. Future work is needed to answer this question too. Third, it will be worth combining investigations that have focused on the retrieval/recognition phase of memory tasks (e.g., Dobbins, 2021; Mill et al., 2016; Võ et al., 2008) and those that focus on the encoding side of the task (e.g., present study; Kucewicz et al., 2018; Papesh et al., 2012) to examine whether similar individual differences account for attention at encoding and attention during memory search. Finally, an interesting extension of this work will be assessing the degree to which intensity of attention and arousal regulation are manipulable. That is, can you encourage people to exert more intensity of attention when they encode information? Does this lead to better memory for that information? Initial work by Ariel and Castel (2014) and Miller et al. (2019) suggests this is the case. Similarly, can you regulate people's arousal in any way? If so, will that improve their memory? Indeed, the present findings beg many questions that are ripe for future investigation.

Conclusion

The present study identified two distinguishable sources of variation in memory ability, both of which were revealed via pupillometry: arousal (dys)regulation and intensity of attention. Specifically, participants who exhibit relatively dysregulated arousal tended to have poorer memory performance, and participants who exhibited greater intensity of attention tended to have better memory performance. These aspects of people constituted distinguishable individual differences, and partially accounted for why people ultimately differed in memory performance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-022-02081-5.

References

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision, 14*, 1–1. https://doi.org/10.1167/14.4.1

Aminihajibashi, S., Hagen, T., Andreassen, O. A., Laeng, B., & Espeseth, T. (2020). The effects of cognitive abilities and task demands on tonic and phasic pupil sizes. *Biological Psychology, 156*, 107945. https://doi.org/10.1016/j.biopsycho.2020.107945

Aminihajibashi, S., Hagen, T., Foldal, M. D., Laeng, B., & Espeseth, T. (2019). Individual differences in resting-state pupil size: Evidence for association between working memory capacity and pupil size variability. *International Journal of Psychophysiology, 140*, 1–7. https://doi.org/10.1016/j.ijpsycho.2019.03.007

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

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

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja. Accessed 21 Sep 2021.

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews, 42*, 33–84. https://doi.org/10.1016/S0165-0173(03)00143-7

Dobbins, I. G. (2021). Pupil dilation signals recognition salience. *Psychonomic Bulletin & Review, 28*(2), 565–573. https://doi.org/10.3758/s13423-020-01866-w

Dowle, M., & Srinivasan, A. (2018). *Data.table: Extension of 'data.frame'*. Retrieved from https://CRAN.R-project.org/package=data.table. Accessed 21 Sep 2021.

Frank, D., & Kafkas, A. (2021). Expectation-driven novelty effects in episodic memory. *Neurobiology of Learning and Memory*, 107466.

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*, 264–281. https://doi.org/10.1037/xlm0000818

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. https://doi.org/10.1037/a0036306

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

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. (2015a). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology, 110*, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Hopstaken, J. F., Linden, D. van der, Bakker, A. B., Kompier, M. A., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 878–889. https://doi.org/10.1037/xhp0000189

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015b). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology, 52*, 305–315. https://doi.org/10.1111/psyp.12339

Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—a review. *Neuroscience & Biobehavioral Reviews, 71*, 58–82. https://doi.org/10.1016/j.neubiorev.2016.08.020

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences, 24*, 466–480. https://doi.org/10.1016/j.tics.2020.03.005

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron, 89*, 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology, 64*, 1971–1989. https://doi.org/10.1080/17470218.2011.588335

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition, 180*, 123–134.

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

Kahneman, D., & Peavler, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology*, *79*, 312–318. https://psycnet.apa.org/doi/https://doi.org/10.1037/h0026912

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336–358. https://doi.org/10.1037/0278-7393.26.2.336

Kleiman, E. (2017). *EMAtools: Data management tools for real-time monitoring/ecological momentary assessment data*. https://CRAN.R-project.org/package=EMAtools. Accessed 21 Sep 2021.

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Vratislav, F., & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports, 8*, 1–7. https://doi.org/10.1038/s41598-018-23197-6

Kursawe, M. A., & Zimmer, H. D. (2015). Costs of storing colour and complex shape in visual working memory: Insights from pupil size and slow waves. *Acta Psychologica, 158*, 67–77. https://doi.org/10.1016/j.actpsy.2015.04.004

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Madore, K. P., Khazenzon, A. M., Backes, C. W., Jiang, J., Uncapher, M. R., Norcia, A. M., & Wagner, A. D. (2020). Memory failure predicted by attention lapsing and media multitasking. *Nature, 587*, 87–91. https://doi.org/10.1038/s41586-020-2870-z

Massar, S. A., Lim, J., Sasmita, K., & Chee, M. W. (2016). Rewards boost sustained attention through higher effort: A value-based decision making approach. *Biological Psychology, 120*, 21–27. https://doi.org/10.1016/j.biopsycho.2016.07.019

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary responses to words that convey a sense of brightness or darkness. *Psychological Science, 28*(8), 1116–1124.

Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience, 8*, 1063. https://doi.org/10.3389/fnhum.2014.01063

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., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language, 104*, 25–42. https://doi.org/10.1016/j.jml.2018.09.005
Miller, A. L., & Unsworth, N. (2018). Individual differences in working memory capacity and search efficiency. *Memory & Cognition, 46*, 1149–1163. https://doi.org/10.3758/s13421-018-0827-3
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*, 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*, 104276. https://doi.org/10.1016/j.jml.2021.104276
Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision, 13*, 11. https://doi.org/10.1167/13.2.11
Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology, 83*, 56–64. https://doi.org/10.1016/j.ijpsycho.2011.10.002
Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*, 559–566. https://doi.org/10.1111/j.1469-8986.1974.tb01114.x
Peckham, A. D., & Johnson, S. L. (2016). Spontaneous eye-blink rate as an index of reward responsivity: Validation and links to bipolar disorder. *Clinical Psychological Science, 4*, 451–463. https://doi.org/10.1177/2167702615594999
R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 21 Sep 2021.
Robison, M. K., & Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception, & Psychophysics, 82*, 3273–3290. https://doi.org/10.3758/s13414-020-02077-0
Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics, 81*, 407–419. https://doi.org/10.3758/s13414-018-1618-4
Robison, M. K., & Brewer, G. A. (In press). Individual differences in working memory capacity, attention control, fluid intelligence, and pupillary measures of arousal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*, 211. https://doi.org/10.1037/0096-3445.126.3.211
Siegle, G. J., Steinhauer, S. R., Stenger, V. A., Konecky, R., & Carter, C. S. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage, 20*, 114–124. https://doi.org/10.1016/S1053-8119(03)00298-2
Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science, 21*, 786–789. https://doi.org/10.1177/0956797610368063
Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1532–1539. https://doi.org/10.1037/a0024852
Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology, 91,* 109–123. https://doi.org/10.1016/j.cogpsych.2016.10.001
Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*(1), 79–139. https://doi.org/10.1037/bul0000176
Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*, 242–254. https://doi.org/10.3758/s13421-012-0261-x
Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132. https://doi.org/10.1037/0033-295X.114.1.104
Unsworth, N. & Miller, A. L. (In press). Individual differences in the intensity and consistency of attention. *Current Directions in Psychological Science*. https://doi.org/10.1177/09637214211030266
Unsworth, N., & Miller, A. L. (2021). Encoding dynamics in free recall: Examining attention allocation with pupillometry. *Memory & Cognition, 49*, 90–111. https://doi.org/10.3758/s13421-020-01077-7
Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Individual differences in lapses of sustained attention: Oculometric indicators of intrinsic alertness. *Journal of Experimental Psychology: Human Perception and Performance, 46*(6), 569–592.https://doi.org/10.1037/xhp0000734
Unsworth, N., Miller, A. L., & Robison, M. K. (2021). Is working memory capacity related to baseline pupil diameter? *Psychonomic Bulletin & Review, 28*, 228–237. https://doi.org/10.3758/s13423-020-01817-5
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*, 757–765. https://doi.org/10.3758/s13423-014-0747-6
Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience, 16*, 601–615. https://doi.org/10.3758/s13415-016-0417-4
Unsworth, N., & Robison, M. K. (2017a). A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review, 24*, 1282–1311. https://doi.org/10.3758/s13423-016-1220-5
Unsworth, N., & Robison, M. K. (2017b). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1962–1987. https://doi.org/10.1037/xlm0000421
Unsworth, N., & Robison, M. K. (2018). Tracking working memory maintenance with pupillometry. *Attention, Perception, & Psychophysics, 80*, 461–484. https://doi.org/10.3758/s13414-017-1455-x
Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392–406. https://doi.org/10.1016/j.jml.2010.02.001
Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *Quarterly Journal of Experimental Psychology, 64*, 447–466. https://doi.org/10.1080/17470218.2010.505292
Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *Journal of Neuroscience, 35*, 7866–7877.
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.

Wainstein, G., Rojas-Líbano, D., Crossley, N., Carrasco, X., Aboitiz, F., & Ossandón, T. (2017). Pupil size tracks attentional performance in attention-deficit/hyperactivity disorder. *Scientific Reports*, *7*, 1–9. Retrieved from https://doi.org/10.1038/s41598-017-08246-w

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse. Accessed 21 Sep 2021.

Wilke, C. O. (2019). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=cowplot. Accessed 21 Sep 2021.

Open Practices Statement To make the present research as transparently reported as possible, we have posted all aggregated datasets and analysis script on the Open Science Framework at the following URL: https://osf.io/275em/. Raw eye-tracking and response data are available upon request. Interested readers are encouraged to contact the corresponding author with questions regarding the analyses.