Received: 5 January 2022 | Revised: 19 May 2022 | Accepted: 20 May 2022

DOI: 10.1111/psyp.14124

PSYCHOPHYSIOLOGY SPR WILEY

ORIGINAL ARTICLE

An examination of relations between baseline pupil measures and cognitive abilities

Matthew K. Robison[1] | Joseph T. Coyne[2] | Ciara Sibley[2] | Noelle L. Brown[2] | Brittany Neilson[2] | Cyrus Foroughi[2]

[1]Department of Psychology, University of Texas at Arlington, Arlington, Texas, USA

[2]U.S. Naval Research Laboratory, Washington, District of Columbia, USA

Correspondence
Matthew K. Robison, University of Texas at Arlington, 501 S. Nedderman Drive, Box 19528, Arlington, TX 76019, USA.
Email: matthew.robison@uta.edu

Abstract

Examining individual differences in pupil size and pupillary dynamics have revealed important insights into the nature of individual differences in cognitive abilities like working memory capacity, long-term memory, attention control, and fluid intelligence. These findings are often tied to the locus coeruleus-norepinephrine (LC-NE) system, as this system has a tight temporal correlation with pupil diameter. Some recent research has demonstrated positive correlations between resting pupil size and cognitive ability, specifically fluid intelligence. The present study attempted to replicate such relations. Across three studies, a large sample of participants (N = 845) completed batteries of cognitive ability measures and measures of resting pupil size and pupillary hippus (fluctuations in pupil diameter). The cognitive measures comprised tasks previously used to measure attention control, visual short-term memory capacity, fluid intelligence, working memory capacity, and visuospatial ability. At the factor level, cognitive ability and pupil size correlated near zero. We did observe some limited evidence for a negative correlation between resting pupillary hippus and cognitive ability. Given the null findings in the present data, we encourage further replication of relations between resting pupil measures and cognitive abilities before making any strong theoretical conclusions about such relations.

KEYWORDS

attention control, fluid intelligence, individual differences, pupillometry, working memory

1 | INTRODUCTION

Although pupil size has been used as a psychophysiological tool in cognitive psychology for many decades (Beatty, 1982; Hess & Polt, 1964; Kahneman & Beatty, 1966), we are continuing to discern exactly what pupil dilation and individual differences therein might reveal about the neurophysiology of cognition. Research has demonstrated a tight temporal linkage between pupil diameter and both tonic and phasic firing of locus coeruleus (LC) neurons, which release most of the norepinephrine (NE) in the brain (Joshi et al., 2016; Rajkowski, 1993; Varazzani et al., 2015). Based on such findings, researchers have been using pupil diameter to theorize about the role of the LC for cognitive functions like vigilance, attention, working memory, long-term memory, and fluid intelligence (Aminihajibashi et al., 2019, 2020; Aston-Jones

& Cohen, 2005; Mäki-Marttunen et al., 2018; Sara, 2009; Tsukahara et al., 2016; Tsukahara & Engle, 2021; Unsworth & Robison, 2017a). Pupillary dynamics have also been leveraged to understand cognitive development during childhood (e.g., Chatham et al., 2009) and cognitive aging/Alzheimer's disease (e.g., Kawasaki et al., 2020; Mather & Harley, 2016). Thus, pupillometry has proven to be a valuable tool to understand the factors underlying cognition in a variety of domains.

The present study focuses on measures of baseline/resting pupil size and intra- and interindividual variability therein, as opposed to evoked pupillary responses within a task context. One particularly intriguing finding is that resting pupil size may correlate with cognitive ability (Heitz et al., 2008; Tsukahara et al., 2016; Tsukahara & Engle, 2021). Specifically, several studies have shown that resting pupil size positively correlates with individual differences in fluid intelligence, working memory capacity, and attention control abilities, with particularly strong relations between fluid intelligence and pupil size (Heitz et al., 2008; Tsukahara et al., 2016; Tsukahara & Engle, 2021). Tsukahara et al. have theorized that the larger pupil sizes among relatively high-intelligence individuals might indicate better functional resting-state coupling between the LC and cortical networks that implement higher-order cognitive functions. Pupillometry is a relatively non-invasive, inexpensive, and portable psychophysiological tool that can be collected on most if not all commercially available eye trackers. If it is indeed the case that resting pupil size correlates with higher-order cognitive abilities like fluid intelligence, this would open doors for future research using this tool as a preliminary step in understanding the role of brain systems like the LC in normal and disordered cognitive development in childhood, individual differences in cognition, and normal and disordered cognitive aging. Indeed, this concept has received attention in the popular press. Tsukahara et al. (2021) recently wrote an editorial publicizing this finding in *Scientific American* magazine, and it has been featured in *Discover* magazine (Learn, 2021).

While the findings from Tsukahara et al. are exciting in their potential, several recent studies have not observed such relations. For example, Unsworth et al. (2019) did not observe significant correlations between resting pupil size and working memory capacity or attention control, and a recent meta-analysis of the correlation between working memory capacity and pupil size estimated the correlation to be quite low and inconsistent across testing sites (Unsworth et al., 2021a). However, these studies did not include measures of fluid intelligence. There are several other studies that have failed to find correlations between pupil size and cognitive abilities. For example, Robison and Brewer (2022) did not observe a significant correlation between resting pupil size and fluid intelligence, working memory capacity, or attention control in a large sample of young adults. Aminihajibashi et al. (2020) also did not observe correlations between pupil size and fluid intelligence or working memory capacity. However, it should be noted that Aminihajibashi et al. (2020)'s pupil size measure was collected during a task, not during a resting measurement.

There are several reasons why these studies may have failed to find correlations between cognitive abilities and pupil size. The most common way to collect resting pupil size is to have participants fixate on a mark centered on a computer screen while an eye-tracker records their pupil diameter. This allows for careful control over the amount of light hitting the eye. But Tsukahara and Engle (2021) indicated several factors worth considering when measuring resting pupil size, especially in relation to individual differences. Generally, they demonstrate that experimental conditions can significantly impact both average pupil size measures and the amount of interindividual variability in the measures. In individual-differences research, it is beneficial to have a large amount of interindividual variation in measures, as range restriction can impact one's ability to observe correlations. Therefore, Tsukahara and Engle (2021) examined whether several factors moderated the correlation between pupil size and cognitive ability. In their first study, a re-analysis of Tsukahara et al. (2016), Tsukahara and Engle compared pupil size from a gray vs. a white background fixation screen. This study showed two important patterns: (1) There was more interindividual variability in pupil size against a gray background and (2) correlations between pupil size and cognitive abilities were significantly stronger with a gray background. In their second study, they carefully manipulated the lighting setting in the experimental room (lights on vs. lights off), the brightness of the computer monitor (bright vs. dim), and the color of the background screen (white vs. black). Tsukahara and Engle (2021) demonstrated that correlations between cognitive abilities and resting pupil measures are strongest under conditions that maximize interindividual variability, namely black backgrounds, lights off, and dim monitor setting. This was an important methodological contribution. However, they also showed that the relation between fluid intelligence and pupil size is rather robust, as there was a significant correlation between pupil size and fluid intelligence under all lighting/brightness/background color settings. Further, they showed that there was a significant positive relation between a general cognitive ability factor comprising measures of fluid intelligence, working memory, and attention control, and a factor representing the common variance among all pupil size measures ($r = .28$).

Given the recent interest in the pupil size-cognition relation, as well as the methodological details described by

Tsukahara and Engle (2021) we believed it was important to attempt an independent conceptual replication with some overlapping and some non-overlapping measures of cognitive ability and multiple measures of resting pupil size against different background illumination settings. We sampled from a large and diverse population (U.S. military officers and enlisted service members). The overarching goal of the present set of studies was to replicate the finding of a positive correlation between cognitive ability, broadly measured, and resting pupil size. We measured pupil size with three fixation screens: black, gray, and white backgrounds. Because of the group-administration experimental setting, we could not also manipulate monitor brightness or room lighting. However, as will be detailed in the Results below, there were large effects of fixation screen background color on both average pupil size measurements and the amount of interindividual variability across participants. Further, measures of central tendency and spread were similar to those observed by Tsukahara and Engle (2021), so we felt confident that we had an adequate moderation of pupil size with this single manipulation.

Finally, in addition to examining individual differences in average pupil size, we also examined individual differences in pupillary hippus, the degree to which an individual's pupil fluctuates from moment to moment, and it can be computed in several different ways (McLaren et al., 1992). To maintain consistency with prior work, we used the intraindividual standard deviation of pupil size during the baseline fixation window as our measure of pupillary hippus. Prior studies have shown that trial-to-trial pupillary hippus within a task correlates with cognitive abilities like working memory capacity (Aminihajibashi et al., 2020; Robison & Brewer, 2020; Robison & Unsworth, 2019; Unsworth & Robison, 2015, 2017a), attention control (Robison & Brewer, 2022; Unsworth & Robison, 2017a), and long-term memory (Madore et al., 2020). Regarding pre-experimental/resting pupillary hippus, there is less data, but one study has shown a positive correlation with working memory capacity (i.e., greater pupillary hippus—higher capacity; Aminihajibashi et al., 2019), and one study has shown negative correlations between resting pupillary hippus and attention control and fluid intelligence (i.e., greater hippus—lower ability; Robison & Brewer, 2022), and some studies have shown null correlations between resting pupillary hippus and cognitive abilities (Tsukahara & Engle, 2021; Unsworth et al., 2019). Therefore, our goal was to bring more data to the question of whether baseline pupil measures (i.e., pupil size and pupillary hippus) correlate with cognitive abilities, and to examine whether indeed these correlations are moderated by environmental factors like background illumination.

2 | STUDY 1

In Study 1, a sample of U.S. military officers completed a battery of cognitive tasks and a resting pupil measurement. The task battery included measures of working memory capacity, attention control, and fluid intelligence.

3 | METHOD

3.1 | Participants and procedure

A sample of 152 participants completed the study. We had limited demographic data for the participants, but it is summarized in Table 1. All participants were student aviators and participated as part of a larger testing battery. Participants completed, in order, an eye-tracker calibration, a baseline/resting pupil measurement, a fixation accuracy measurement (not analyzed here), a digit span task, a direction orientation task (not analyzed here), an antisaccade task (Hutchison, 2007), an orientation-judgment visual arrays task (Luck & Vogel, 1997; Vogel et al., 2005), a rotation complex span task (Kane et al., 2004), the Raven Advanced Progressive Matrices, the Sustained Attention to Cue Task (SACT; Draheim et al., 2021), and a demographic survey. Here, we present the data from the baseline pupil measurement, antisaccade, visual arrays, mental counters task, digit span, and SACT. Sessions were administered once in the morning starting at 9:00 a.m. and once in the afternoon starting at 1:00 p.m. Participants completed the sessions in a group setting. There were 15 stations in the experimental room, but a maximum of 8 participants completed the study at one time. The experimental room was windowless but well-lit. Participants wore headphones during the tasks, as some included auditory stimulation (e.g., antisaccade). The experimental protocol was approved by the Institutional Review Board of the U.S. Naval Research Laboratory.

3.1.1 | Digit span

On each trial, a sequence of 3, 5, 7, or 9 single digits appeared on the screen one at a time. At the end of this list, participants were instructed to report the digits as they had appeared in the correct serial order. A digit was scored as correct if it was reported in the correct serial position, and the score for a trial was the number of digits correctly reported. Participants completed 12 lists (3 lists of each set size). The dependent variable was the total number of correctly reported digits (max score = 72).

TABLE 1 Demographic statistics for sample populations

	Study 1	Study 2	Study 3
Age (SD)	23.66 (2.10)	23.28 (3.18)	21.65 (3.73)
Age range	21–31	18–38	18–38
Male	81%	86%	82%
Female	16%	13%	15%
Asian		3%	3%
Black or African American		8%	12%
Hispanic or Latino		9%	15%
Native American		0%	1%
Native Hawaiian or Pacific Islander		1%	1%
Multiracial		12%	14%
White		65%	52%
Some high school		0%	0%
High school diploma		17%	51%
Some college		10%	21%
College degree		65%	23%
Graduate school		5%	3%
Aviator	100%	66%	20%
Air-traffic control		33%	6%
Mechanic			35%
Ordnanceman			38%

3.1.2 | Antisaccade

On each trial, a central fixation point appeared for either 1000 or 2000 ms. Then, a 300-ms auditory beep warned participants that the cue was imminent. Then, an asterisk (*) appeared for 300 ms either to the right or left of fixation. Then, a target letter (O or Q) appeared on the opposite side of the screen as the cue, followed by a backward visual mask (##). The mask remained on-screen until participants made a response or until 5 s had elapsed. Participants used the O and Q keys on the keyboard to make their responses. Participants completed 8 practice trials where the target letter appeared for 500 ms before being masked, then 16 practice trials where the target letter appeared for 100 ms before being masked, then 72 scored trials with a 100-ms target duration. The target was equally likely to be an O or a Q and equally likely to appear on the right or left side of the screen. Participants received feedback on every trial regarding accuracy ("correct" in cyan font, "incorrect" in magenta font, or "no response detected" in red font.) The dependent variable was the proportion of correctly-identified letters on the scored trials.

3.1.3 | Visual arrays

Each trial was initiated by the participant pressing the space bar. Then, a 1000-ms fixation cross appeared centered against a gray background. Then, a cue word (BLUE or RED) appeared at the center of the screen for 300 ms. Then, a pattern of 12 oriented rectangles appeared on the screen. The four possible orientations were vertical, horizontal, angled 45° right, and angled 45° left. On each trial, there were either 7 cued-color items and 5 non-cued color items or 5 cued-color and 7 non-cued color items. Participants were instructed to only pay attention to the cued-colored items, as they would only be tested. The items appeared for 250 ms followed by a 900-ms blank delay. Then, the cued-colored items reappeared. One rectangle had a white dot on it. The participant's task was to indicate whether this rectangle was the same orientation or a different orientation as the initial presentation. Participants used the 5 and 6 keys on the number pad to make their responses. The keys were marked "yes" meaning same orientation or "no" meaning different orientation, respectively. Participants completed 3 practice trials with feedback and 80 scored trials without feedback. Trials were equally likely to be red/blue target color, set size 5 or 7 target items, and change/no-change in orientation for the tested item. The dependent variable was proportion correct.[1]

3.1.4 | Rotation span

In this task, participants were presented with sequences that interleaved a judgment about a letter (mirrored or not) with remembering the size and direction of a pointed arrow (Kane et al., 2004). At the end of the sequence, participants reported both the direction and size of the arrows in forward serial order. There were two sequences each of 2, 3, 4, 5, 6, and 7 arrows (12 total lists). The dependent variable was the total number of correctly-reported arrows (i.e., correct arrow in the correct serial position.)

3.1.5 | Raven advanced progressive matrices

On each trial, participants were presented with a 3×3 matrix that formed a patterned grid. The bottom-right piece of the grid was missing, and the participants' task was to select from a set of 8 possible pieces a solution that best completed implicit patterns in the grid. Participants received 3 practice items with solutions explained. Then, they had 10 min to complete 18 items. The dependent variable was a total number of items correctly solved.

3.1.6 | Sustained attention to cue task

Each trial started with a black fixation point against a silver background. Then, a 300-ms auditory tone alerted participants to the start of a trial. A large white circle then appeared on either the right or left side of the screen. Then, after a waiting period of 2–12 s, a white distractor (*) flashed at the center of the screen for 300 ms. A 3×3 array of letters then appeared at the previously cued location for 125 ms. The array contained the letters B, P, D, and R. The target letter was the centermost letter, and the participants' task was to identify that letter. After 125 ms, a mask appeared over the central letter for 1000 ms. The possible responses were then shown on the screen horizontally, and participants clicked the target letter. Participants completed 64 scored trials after 5 practice trials that included feedback. The scored trials did not contain feedback. The dependent variable was the proportion of correctly identified target letters.

[1]It is also common for this task to compute a capacity (k) estimate, which accounts for set size, hit rates, and false alarm rates. The correlation between proportion correct and k in the sample was 0.997.

3.1.7 | Baseline/resting pupil measure

Pupil data were collected via Gazepoint GP3HD eye-trackers sampling at 150 Hz.[2] The eye trackers were mounted to the bottom of the computer monitors. After setting up the eye tracking software, participants fixated on a silver fixation mark for 90 s (30 s against a black background, followed directly by 30 s against a gray background, followed directly by 30 s against a white background). Participants were instructed to keep their eyes on the fixation mark, but that they could blink as needed. Participants wore a pair of glasses with the lenses removed, onto which a fiducial marker with a known diameter was taped. The eye-tracker used this marker to convert pupil diameter measures from pixels to millimeters. From the resting pupil measurement, we computed the mean pupil diameter and pupillary hippus for each participant for each background color. We excluded the first 5 s of each 30-s window to allow for an adjustment of the pupil to the level of background luminance. Missing data due to blinks and off-screen fixations were excluded. Any participant who had more than 40% of samples missing was excluded from the analysis. Participants freely viewed the fixation screen (i.e., did not have head stabilized in a chinrest).

3.2 | Data analysis

We used R software for all our analyses. Data were aggregated and transformed using the *tidyverse* set of packages (Wickham, 2019); plots were generated using the *ggplot2* (Wickham, 2016) and *cowplot* (Wilke, 2020) packages; data were analyzed using the *psych* (Revelle, 2018), *lavaan* (Rosseel, 2012), *lmerTest* (Kuznetsova et al., 2017), and *PairedData* (Champely, 2018) packages. The manuscript was written in R Markdown using the *papaja* (Aust & Barth, 2018) package. Pupil data were preprocessed using custom R scripts. The Gazepoint software automatically scores whether a

[2]Mannaru et al. (2017) manipulated screen luminance (black, gray, and white backgrounds) and task load to demonstrate the pupillary measurement quality of the Gazepoint GP3 Eye Tracker. Classifications of high compared to low task load with black, gray, and white backgrounds were 90.75%, 87.53%, and 86.89%, respectively. This supports the quality of the pupillary data recorded from this eye tracker, such that it can distinguish pupillary changes associated with both luminance and cognitive load. More recently, Cuve et al. (2022) computed several accuracy and precision metrics for this eye tracker, and all metrics supported its use. Most relevant to this study on pupillometry, they also observed significant pupil dilation changes associated with differences in screen luminance (black vs. white backgrounds) that were relatively equivalent with and without a chinrest.

TABLE 2 Descriptive statistics for cognitive ability measures in all studies

Study	Task	*N*	Mean	SD	Skew	Kurtosis	Reliability
Study 1	Digit span	132	59.41	8.40	−0.56	−0.44	0.53
	Antisaccade	123	0.82	0.12	−0.72	−0.68	0.88
	Visual arrays	124	0.70	0.09	−0.07	−0.97	0.77
	Rotation span	130	31.63	8.37	−0.23	−0.16	0.55
	Raven	132	11.38	2.42	−0.55	0.42	0.55
	SACT	128	0.93	0.05	−1.17	1.54	0.80
Study 2	Antisaccade	244	0.83	0.09	−0.79	−0.07	0.83
	Visual arrays	214	0.69	0.09	−0.02	−0.46	0.67
	Digit span	252	57.87	7.52	−0.24	−0.63	0.56
	Mental counters	261	0.66	0.17	−0.26	−0.63	0.85
	SACT	234	0.91	0.08	−1.20	0.83	0.89
Study 3	Antisaccade	347	0.82	0.13	−0.77	−0.46	0.89
	Visual arrays	341	0.62	0.09	0.73	−0.10	0.65
	Terrain orientation (RMSE)	447	93.38	60.57	0.56	−0.37	0.90
	Mental counters	447	16.68	7.54	−0.42	−0.48	0.90
	Raven	428	7.56	3.63	0.11	−0.69	0.87

Abbreviations: SACT, sustained attention to cue task; SD, standard deviation.

pupil sample is valid or not (1 or 0). We used this validity indicator to filter out invalid samples. We used values from the right eye for all analyses. To compute reliability for the pupil measures, we separately computed each dependent measure (e.g., mean pupil diameter) by odd and even samples, then computed a correlation between the measured extracted from the odd and even samples, and finally applied the Spearman-Brown split-half correction to the correlation. For most cognitive tasks, we performed a similar procedure. First, we split trials between odd and even trials, then computed the dependent variable (e.g., proportion correct, total number correct, RMSE, etc.), then correlated those measures, and applied the split-half correction. For the digit span and rotation span tasks, we used Cronbach's alpha on performance at each set size to estimate reliability.

4 | RESULTS AND DISCUSSION

A descriptive summary of performance on each cognitive task is listed in Tables 2 and 3 lists a descriptive summary of the pupil measures, and Table 4 lists correlations among the measures.

Our first set of analyses examined pupil diameter and pupillary hippus against each background screen. Figures 1 and 2 show the distributions of average pupil diameter and pupillary hippus against each background screen as a raincloud plot (Allen et al., 2019). As would be expected, pupil diameter was significantly larger against the black background compared to the gray background ($b = 1.55$, 95% CI = [1.46, 1.64], $p < .001$) and significantly smaller against the white background compared to the gray background ($b = -0.64$, 95% CI = [−0.74, −0.55], $p < .001$). There was significantly more pupillary hippus against the black background compared to the gray background ($b = 0.06$, 95% CI = [0.04, 0.08], $p < .001$), but there was no significant difference in hippus between the gray and white backgrounds ($b = 0.0002$, 95% CI = [−0.02, 0.02], $p = .98$).

As highlighted by Tsukahara and Engle (2021), darker luminance conditions can also produce more interindividual variation in pupil measures. So, we examined that aspect of the data, as well, using the Pitman-Morgan test of variances in paired samples. Compared to the gray background, the black background produced more interindividual variability in both pupil diameter ($t[107] = 9.60$, $p < .001$) and pupillary hippus ($t[105] = 8.10$, $p < .001$). Interestingly, the gray background produced *more* variability in pupil diameter than the white background ($t[106] = 3.50$, $p < .001$) but significantly *less* variability in pupillary hippus compared to the black background ($t[106] = -4.54$, $p < .001$). Our analyses largely replicated the pattern observed by Tsukahara and Engle (2021) that conditions with low environmental luminance produce the most interindividual variability in pupil measures, and thus may be best suited for examinations of individual differences.

TABLE 3 Descriptive statistics for pupil measures in all studies

Study	Measure	*N*	Mean	SD	Skew	Kurtosis	Reliability
Study 1	Black background—Mean	111	4.59	0.85	0.24	0.10	>.99
	Gray background—Mean	109	3.04	0.41	0.05	0.25	>.99
	White background—Mean	109	2.39	0.34	0.32	−0.15	>.99
	Black background—SD	107	0.28	0.14	2.13	5.47	>.99
	Gray background—SD	109	0.21	0.07	0.99	0.69	>.99
	White background—SD	110	0.20	0.10	2.32	6.81	>.99
Study 2	Black background—Mean	263	4.34	0.78	0.24	−0.57	>.99
	Gray background—Mean	263	3.09	0.41	0.16	−0.04	>.99
	White background—Mean	262	2.47	0.37	0.17	−0.32	>.99
	Black background—SD	257	0.24	0.11	1.69	4.07	>.99
	Gray background—SD	258	0.22	0.09	1.16	1.28	>.99
	White background—SD	258	0.19	0.07	0.96	0.65	>.99
Study 3	Black background—Mean	443	4.56	0.87	−0.07	0.04	>.99
	Gray background—Mean	442	3.06	0.47	0.17	0.72	>.99
	White background—Mean	444	2.38	0.36	0.26	0.16	>.99
	Black background—SD	432	0.32	0.23	2.48	7.20	>.99
	Gray background—SD	439	0.26	0.13	1.61	3.33	>.99
	White background—SD	439	0.22	0.09	1.22	1.92	.99

TABLE 4 Zero-order correlations among measures in Study 1

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Pupil—black M	–										
2. Pupil—gray M	**0.68**	–									
3. Pupil—white M	**0.57**	**0.82**	–								
4. Pupil—black SD	0.01	**0.31**	**0.26**	–							
5. Pupil—gray SD	0.05	**0.27**	**0.31**	**0.41**	–						
6. Pupil—white SD	−0.03	**0.19**	**0.42**	**0.19**	**0.64**	–					
7. Digit span	0.01	−0.09	0.02	0.06	−0.03	**−0.23**	–				
8. Antisaccade	0.09	0.07	0.04	−0.04	0.00	**−0.24**	**0.25**	–			
9. Visual arrays	0.07	0.02	−0.08	−0.10	−0.12	**−0.22**	**0.36**	**0.30**	–		
10. Rotation span	0.12	−0.04	0.03	−0.14	−0.08	−0.14	**0.32**	**0.21**	**0.23**	–	
11. Raven	−0.03	**−0.33**	**−0.23**	**−0.25**	−0.14	**−0.25**	**0.38**	0.15	**0.23**	**0.35**	–
12. SACT	0.09	−0.03	−0.08	−0.04	−0.16	−0.16	−0.04	**0.33**	**0.25**	0.08	−0.05

Note: Bolded correlations are significant at $p < .05$.

Abbreviations: M, mean pupil diameter; SACT, sustained attention to cue task; SD, standard deviation of pupil diameter (hippus).

Next, we examined the relations among the pupil measures and the cognitive measures. The correlations between the pupil and cognitive measures are listed in Table 4. Most of the zero-order correlations were weak and nonsignificant. To examine whether the correlations were significantly moderated by background color, we specified linear mixed-effects models with pupil diameter as the dependent measure and background color and each cognitive measure as fixed effects, as well as their interaction. Cognitive and pupil measures were standardized, and background color was sum-to-zero effects coded (black = −1, gray = 0, white = 1). Participants were specified as random effects. If there is a significant association between pupil diameter and a cognitive measure, the model would reveal a significant main effect of the cognitive measure. If the correlation between the cognitive measure is significantly moderated by background color, this would produce a cognitive measure × background color interaction.

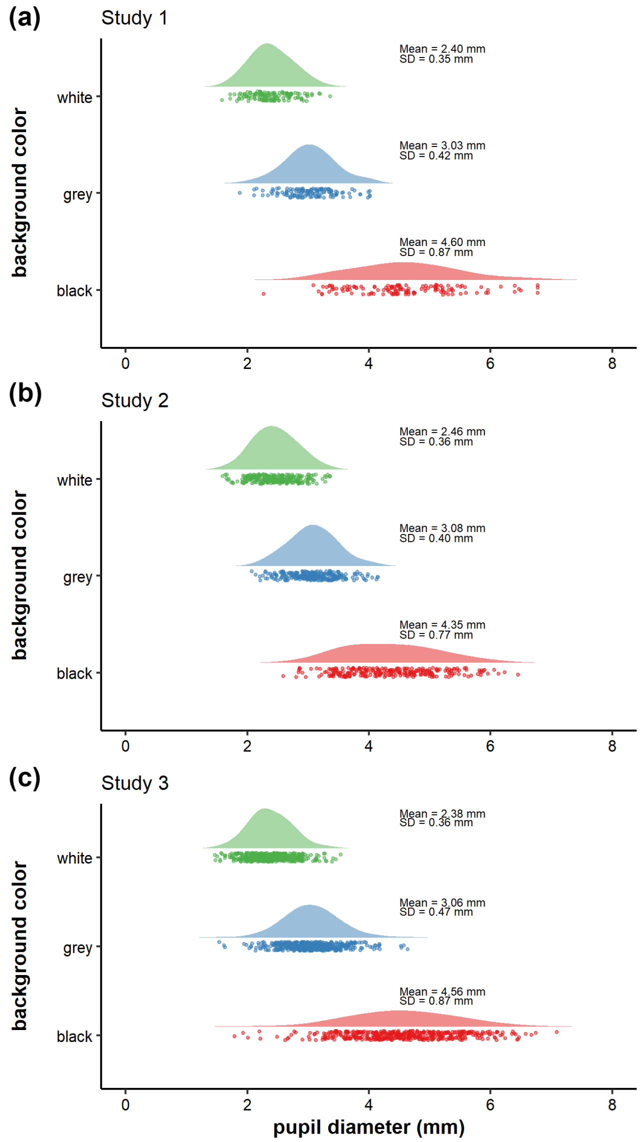

FIGURE 1 Distributions of pupil diameter by background color for Studies 1, 2, and 3

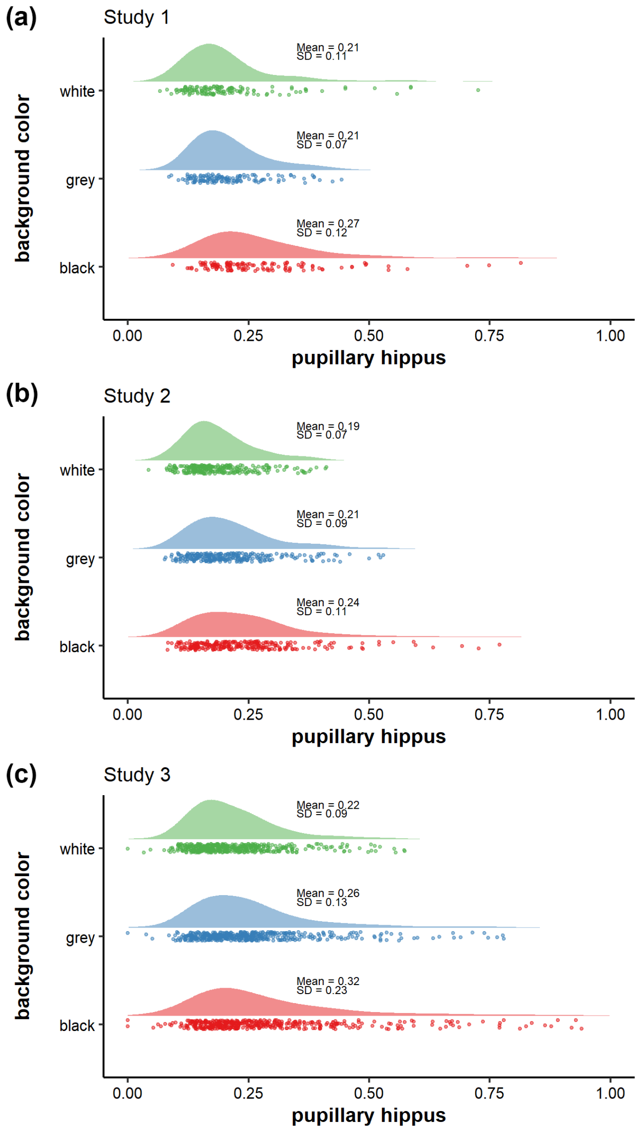

FIGURE 2 Distributions of pupillary hippus by background color for Studies 1, 2, and 3

The models on pupil diameter are summarized in Table 5. There were significant visual arrays × background, Raven × background, and SACT × background interactions, indicating that the correlation between pupil diameter and visual arrays, Raven, and SACT were all moderated by background color. In all three cases, the correlation was more positive against a black background and more negative against a white background. This is also consistent with the moderation demonstrated by Tsukahara and Engle (2021). However, it is worth noting that none of the cognitive measures showed significant positive correlations with pupil diameter, either overall or in the three specific background conditions (see Table 4).

Next, we repeated the above analysis with pupillary hippus as the dependent measure, rather than pupil diameter. The models are summarized in Table 6. Only the correlation between digit span and pupillary hippus was significantly moderated by background color (antisaccade showed a marginal interaction). There was a significant negative main effect of Raven on pupillary hippus and a marginal main effect of visual arrays on pupillary hippus. No other main effects or interactions were significant.

As a final step in the analyses, we examined overall relations between pupil diameter, pupillary hippus, and cognitive ability with z score composites. We averaged standardized scores for the digit span, antisaccade, visual arrays, rotation span, Raven, and SACT and examined this correlation with a z score composite for pupil diameter and pupillary hippus against each background. For the measures against the black background, there was neither a significant correlation between cognitive ability and pupil diameter ($r[132] = .09$, $p = .31$) nor pupillary hippus

TABLE 5 Summaries of linear mixed effect models on pupil diameter in Study 1

Effect	B	SE	p
Digit span	−0.01	0.09	.96
Digit span × background	0.04	0.03	.15
Antisaccade	0.10	0.09	.26
Antisaccade × background	−0.02	0.02	.45
Visual arrays	−0.02	0.08	.77
Visual arrays × background	**−0.06**	**0.02**	**.01**
Rotation span	0.05	0.09	.55
Rotation span × background	−0.04	0.02	.08
Raven	−0.15	0.08	.07
Raven × background	**−0.07**	**0.03**	**.005**
SACT	−0.02	0.10	.87
SACT × background	**−0.11**	**0.03**	**<.001**

Abbreviations: *B*, standardized regression coefficient; SACT, Sustained Attention to Cue Task; *SE*, standard error of estimate.
Bolded values are significant at $p < .05$.

TABLE 6 Summaries of linear mixed effect models on pupillary hippus in Study 1

Effect	B	SE	p
Digit span	−0.04	0.06	.44
Digit span × background	**−0.07**	**0.03**	**.006**
Antisaccade	−0.03	0.06	.58
Antisaccade × background	**−0.05**	**0.02**	**.05**
Visual arrays	**−0.10**	**0.05**	**.05**
Visual arrays × background	−0.03	0.02	.18
Rotation span	−0.09	0.05	.10
Rotation span × background	−0.01	0.02	.67
Raven	**−0.14**	**0.05**	**.006**
Raven × background	0.01	0.03	.61
SACT	−0.08	0.06	.17
SACT × background	−0.04	0.03	.17

Abbreviations: *B*, standardized regression coefficient; SACT, sustained attention to cue task; *SE*, standard error of estimate.
Bolded values are significant at $p < .05$.

($r[127] = -.13$, $p = .13$). For the measures against the gray background, there was neither a significant correlation between cognitive ability and pupil diameter ($r[130] = -.12$, $p = .18$), nor with pupillary hippus ($r[130] = -.13$, $p = .13$). For the measures against the white background, there was not a correlation between cognitive ability and pupil diameter ($r[130] = -.07$, $p = .41$), but there was a significant negative correlation with pupillary hippus ($r[130] = -.33$, $p < .001$). The correlation between the pupil diameter *z* score composite (averaged across all backgrounds) and the cognitive ability composite was small and non-significant ($r[132] = -.02$, $p = .78$), but the correlation between the pupillary hippus *z* score composite and the cognitive ability composite was significantly negative ($r[132] = -.25$, $p = .003$). That is, people who had more fluctuations in pupil diameter during the resting measurement tended to exhibit poorer cognitive performance overall.

In Study 1, relations between cognitive ability and resting pupil measures were largely small, although there was a significant negative correlation between average pupillary hippus and a cognitive ability *z* score composite. While the sample was reasonably large to detect correlations, it was restricted to student aviators, all of whom had college degrees. Studies 2 and 3 expanded the sample and broadened the population to include enlisted military members who are only required to have a high school diploma. The goal was to increase interindividual variability—both in cognitive ability and potentially resting pupil measures—in the sample and to increase statistical power to detect even small correlations.

5 | STUDY 2

In Study 2, a sample of U.S. military officers and enlisted service members completed a battery of cognitive tasks in addition to a resting pupil measure. The sample in Study 2 was expanded in size and included participants from an additional population in the U.S. military. Whereas Study 1 only included aviators, Study 2 included both student aviators and student air-traffic controllers.

5.1 | Method

5.1.1 | Participants and procedure

A sample of 299 participants completed the study. We had to exclude 37 participants because of duplicated participant identification numbers or for having missing data for more than four tasks. Therefore, the final sample included 262 participants. Table 1 summarizes demographic data for the sample. Sessions lasted 2 h. Participants completed, in order, an eye-tracker calibration, a baseline/resting pupil measurement, a direction orientation task, a mental counters task (Alderton et al., 1997), the same antisaccade task as Study 1, the same visual arrays task as Study 1, the same SACT as Study 1, the same digit span task as Study 1, and a demographic survey. Here, we present the data from the baseline pupil measurement, antisaccade, visual arrays, mental counters task, digit span, and SACT. The direction orientation task included two separate versions across participants and was not analyzed here.

Sessions were administered once in the morning starting at 9:00 a.m. and once in the afternoon starting at 1:00 p.m. Participants completed the sessions in a group setting in the same windowless, well-lit room as in Study 1. Participants wore headphones during the tasks, as some included auditory stimulation (e.g., antisaccade). The experimental protocol was approved by the Institutional Review Board of the U.S. Naval Research Laboratory.

5.1.2 | Baseline pupil measurement

See Study 1.

5.1.3 | Antisaccade

See Study 1.

5.1.4 | Visual arrays

See Study 1.

5.1.5 | Mental counters

The mental counters task is used in military selection as an indicator of working memory capacity. As previously described by Alderton et al. (1997), this task requires participants to maintain three values in memory at a time. On each trial, three horizontal lines appeared, and the instructions indicated that each line started with the value 5. Boxes then appeared above or below each line in sequence. If a box appeared above a line, participants added 1 to that line's running total. If a box appeared below the line, the participant subtracted 1 from that line's total. After receiving instructions and practice trials with feedback, participants completed 32 experimental trials. The trial was scored as correct if participants correctly reported all 3 finishing values at the end of a trial.

5.1.6 | Digit span

See Study 1.

5.1.7 | SACT

See Study 1.

5.2 | Data analysis

Data analysis procedures were identical to Study 1.

6 | RESULTS AND DISCUSSION

Table 2 shows a summary of performance on the cognitive measures. The cognitive measures were rather normally distributed with low values for kurtosis and skewness. There was also a rather broad range of ability in each measure, and the low kurtosis values indicate participants were not on the ceiling or on floor. The

TABLE 7 Correlations among measures in Study 2

Measure	1	2	3	4	5	6	7	8	9	10
1. Pupil—black M	–									
2. Pupil—gray M	**0.69**	–								
3. Pupil—white M	**0.51**	**0.82**	–							
4. Pupil—black SD	**0.26**	**0.21**	**0.22**	–						
5. Pupil—gray SD	−0.02	0.08	−0.03	**0.25**	–					
6. Pupil—white SD	0.10	**0.18**	0.12	**0.38**	**0.48**	–				
7. Antisaccade	−0.01	−0.07	−0.05	−0.01	0.00	−0.06	–			
8. Visual arrays	0.05	0.04	0.09	−0.07	**−.14**	−0.11	**0.32**	–		
9. Digit span	0.00	−0.02	−0.03	−0.09	0.03	0.03	**0.19**	**0.23**	–	
10. Mental counters	0.03	0.04	0.03	0.00	0.03	−0.05	**0.23**	**0.39**	**0.23**	–
11. SACT	0.08	0.01	−0.03	**−0.14**	−0.02	−0.09	**0.23**	**0.24**	**0.28**	**0.18**

Note: Bolded correlations are significant at $p < .05$.

Abbreviations: M, mean pupil diameter; SACT, Sustained Attention to Cue Task; SD, standard deviation of pupil diameter (hippus).

cognitive measures were all significantly intercorrelated (see Table 7).

Our next set of analyses examined the baseline/resting pupil size measurement. Figure 1 shows the distributions of pupil diameter against each background and Figure 2 shows the distributions for pupillary hippus. The results of these analyses all replicated Study 1. Descriptive statistics are listed in Table 3, and correlations among the measures are listed in Table 7. Pupil diameters were significantly smaller against a white background screen (b = −0.61, 95% CI = [−0.68, −0.55], p < .001) and significantly larger against a black background screen (b = 1.29, 95% CI = [1.22, 1.35], p < .001) compared to a gray background screen.

There was also significantly more variability in pupil diameter against the black screen compared to the gray screen (t[257] = 15.53, p < .001), and significantly more variability with the gray screen compared to the white screen (t[255] = −2.72, p = .007). This replicates Study 1 and Tsukahara and Engle (2021). The black background also produced significantly greater pupillary hippus compared to the gray screen (b = 0.03, 95% CI = [0.01, 0.04], p < .001), and the white screen produced significantly less pupillary hippus than the gray screen (b = −0.03, 95% CI = [−0.04, −0.01], p < .001). Interestingly, there was more interindividual variability in pupillary hippus across the different background screens, as well. Specifically, there was more variability in pupillary hippus against the black screen compared to the gray screen (t[249] = 4.43, p < .001), and there was more variability in pupillary hippus against the gray screen compared to the white screen (t[248] = −5.12, p < .001).

We next examined correlations between the pupil measures and cognitive ability measures. Overall, the correlations were mostly small and most were not significant, similar to Study 1. As shown in Table 7, there were no significant correlations between pupil diameter and cognitive measures, and there were only two significant correlations between pupillary hippus and cognitive measures. SACT accuracy negatively correlated with pupillary hippus against a black background, and visual arrays accuracy negatively correlated with pupillary hippus against a gray background.

To examine whether the correlations were significantly moderated by background screen color, we specified the same set of linear mixed models as in Study 1. The models on mean pupil diameter are summarized in Table 8. Only one measure showed a performance × background color interaction. The correlation between SACT accuracy and pupil diameter was stronger (more negative) against the black background compared to white and gray backgrounds. No other main effects or interactions were significant. The models on pupillary hippus are summarized in Table 9. There was only one significant effect: a negative main effect of visual array accuracy on pupillary hippus.

As our final analysis in Study 2, we again specified factors for cognitive ability, pupil diameter, and pupillary hippus. However, instead of creating z score composites, we used factor analysis to capture shared variance in the measures. The sample size in Study 1 was not sufficient for stable factor-analytic estimates, but the sample size in Study 2 was sufficiently large for this analysis (Kretzschmar & Gignac, 2019). Specifically, we allowed the five cognitive measures (digit span, antisaccade, visual arrays, mental counters, and SACT) to load onto a Cognitive Ability factor, measures of mean pupil diameter from the black, gray, and white screens to load onto a Pupil Diameter factor, and measures of

TABLE 8 Summaries of linear mixed effect models on pupil diameter in Study 2

Effect	B	SE	p
Antisaccade	−0.04	0.06	.52
Antisaccade × background	−0.02	0.03	.51
Visual arrays	0.06	0.06	.30
Visual arrays × background	0.01	0.03	.66
Digit span	−0.02	0.06	.77
Digit span × background	−0.01	0.03	.62
Mental counters	0.04	0.06	.50
Mental counters × background	0.00	0.03	.99
SACT	0.04	0.07	.57
SACT × background	**−0.07**	**0.03**	**.01**

Abbreviations: B, standardized regression coefficient; SACT, sustained attention to cue task; SE, standard error of estimate.
Bolded values are significant at p < .05.

TABLE 9 Summaries of linear mixed effect models on pupillary hippus in Study 2

Effect	B	SE	p
Antisaccade	−0.02	0.03	.50
Antisaccade × background	−0.01	0.02	.46
Visual arrays	**−0.06**	**0.03**	**.04**
Visual arrays × background	−0.01	0.02	.57
Digit span	0.00	0.03	.89
Digit span × background	0.03	0.02	.14
Mental counters	0.00	0.03	.88
Mental counters × background	−0.01	0.02	.47
SACT	−0.06	0.03	.08
SACT × background	0.01	0.02	.77

Abbreviations: B, standardized regression coefficient; SACT, sustained attention to cue task; SE, standard error of estimate.
Bolded values are significant at p < .05.

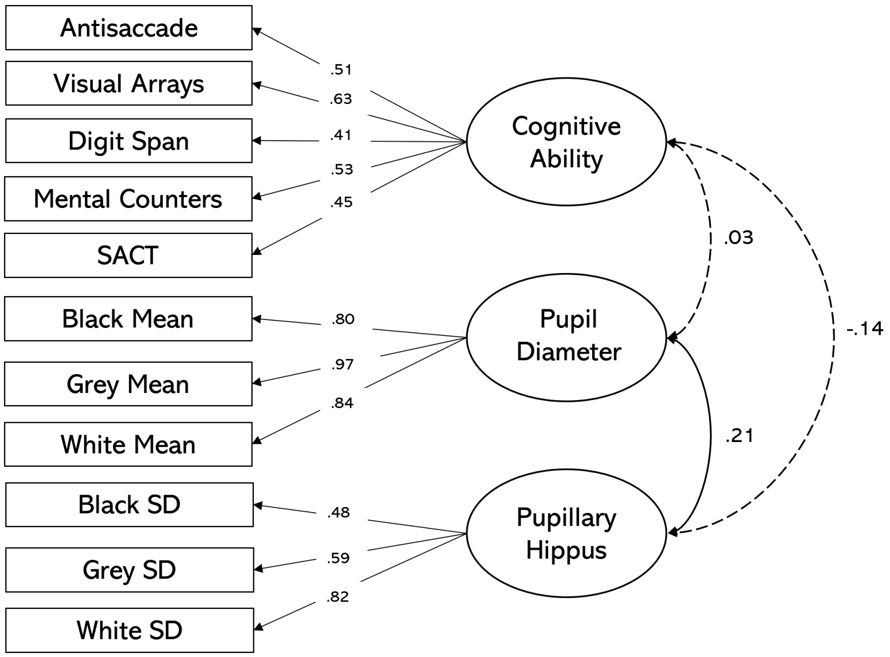

FIGURE 3 Latent factor analysis of the relations among Cognitive Ability, Pupil Diameter, and Pupillary Hippus in Study 2. Solid lines indicate significant paths at $p < .05$, dashed lines indicate non-significant paths

pupillary hippus from the black, gray, and white screens to load onto a Pupillary Hippus factor. The model is depicted in Figure 3.[3] The model fit the data acceptably $\chi^2(42) = 98.98$, CFI = 0.92, TLI = 0.89, RMSEA = 0.07, SRMR = 0.08).[4] All of the measures loaded significantly onto their respective factors. However, there were no significant correlations between cognitive ability and either pupil diameter or pupillary hippus. The cognitive ability factor did not significantly correlate with pupil diameter against the black ($r[255] = .05$, $p = .41$), gray ($r[255] = .01$, $p = .83$), or white background ($r[255] = .02$, $p = .74$), nor with pupillary hippus against the black ($r[252] = -.07$, $p = .24$), gray($r[252] = -.04$, $p = .50$), or white background ($r[252] = -.08$, $p = .18$), individually.

Overall, correlations among cognitive measures and resting pupil measures were quite small and mostly non-significant in Study 2, similar to Study 1. Despite a larger and broader sample, we did not observe any strong evidence for an association between either pupil diameter or pupillary hippus and cognitive ability, regardless of the color of the background screen against which the pupil was measured.

[3]Allowing the mean pupil diameter measures to load freely led to a Heywood case, with pupil diameter against the gray background having a standardized loading greater than 1. Therefore, we had to specify this loading to equal 1.

[4]CFI = confirmatory fit index, TLI = Tucker-Lewis index, RMSEA = root mean squared error of approximation, SRMR = standardized root mean residual. CFI and TLI values closer to 1.0 indicate better fit, and RMSEA and SRMR values closer to 0 indicate better fit.

7 | STUDY 3

Study 3 replicated and extended some components of Studies 1 and 2. First, the sample size was nearly doubled from Study 2, giving us even more power to detect small effects. Second, the Raven Advanced Progressive Matrices were added back to the test battery. Third, we dropped the digit span task, which had relatively poor reliability and loaded only weakly onto the cognitive ability factor in Study 2. Finally, the sample population was diversified to include student mechanics and ordnancemen in addition to student aviators and air-traffic controllers. While the student air traffic controllers were an enlisted population, the score required on the Armed Services Vocational Aptitude Battery (ASVAB) is higher than most other military occupations. The mechanics and ordnancemen had lower ASVAB scores required to be admitted. Our hope was that we would get a broader sample of both cognitive abilities and pupil sizes by widening the population from which we sampled.

8 | METHOD

8.1 | Participants and procedure

A sample of 585 United States military officers and enlisted service members completed the study. We had to exclude 138 for a number of reasons. A large group of participants was excluded because one station had a lower monitor brightness setting than all other stations, which were identical. This station led to significantly smaller pupil measures and thus these participants were excluded. Other participants were excluded because they were either missing data for more than four tasks or were labeled as multivariate outliers. Therefore, the final sample included 447 participants.[5,6] Sessions lasted 2 h. Participants completed, in order, a demographic survey, an eye-tracker calibration, a baseline/resting pupil measurement, the same mental counters task as Study 2, a terrain orientation task (Ostoin, 2007), a color-word Stroop task (Stroop, 1935), an arrow flanker task (Eriksen & Eriksen, 1974), a Simon task (Simon, 1990), the same orientation-judgment visual arrays task as Studies 1 and 2, the same antisaccade task as Studies 1 and 2, the same Raven Advanced Progressive Matrices task as Study 1, and the SynWin task (Elsmore, 1994; Hambrick et al., 2010). Here, we present the data from the baseline pupil measurement, mental counters, terrain orientation, visual arrays, antisaccade, and Raven tasks. The Stroop, flanker, and Simon tasks were novel iterations of those tasks collected as part of a separate project in combination with SynWin (manuscript forthcoming). Sessions were administered once in the morning starting at 9:00 a.m. and once in the afternoon starting at 1:00 p.m. Participants completed the sessions in a group setting in the same windowless, well-lit room as Studies 1 and 2. Participants wore headphones during the tasks, as some included auditory stimulation (e.g., antisaccade, flanker). The experimental protocol was approved by the Institutional Review Board of the U.S. Naval Research Laboratory.

8.1.1 | Baseline pupil measurement

See Study 1.

[5]To assess inter-station reliability in pupil diameter and pupillary hippus, we had a separate group of 57 participants complete the pre-experimental pupil measurement at all 8 stations. The measurements were highly reliable across stations (white background mean: $\alpha = 0.95$, gray background mean: $\alpha = 0.97$, black background mean: $\alpha = 0.97$, white background SD: $\alpha = 0.75$, gray background SD: $\alpha = 0.62$, black background SD: $\alpha = 0.83$). Next, we entered each pupil measure (mean and SD against each background) to a model with a station as a predictor and saved the associated residual values. All these residual values correlated with the original values at 0.98 or above. Finally, we ran the latent variable analyses entering the residuals instead of original values, and the results were nearly identical. Therefore, we do not believe our results were affected by any potential inter-station differences, other than the one noted above.

[6]We also ensured that the excluded participants did not differ from the full sample in demographic makeup or cognitive ability.

8.1.2 | Terrain Orientation Task

The terrain orientation task is a newly developed task that requires visuospatial abilities to complete. The task is a modified version of the direction orientation task that is currently used in military selection (see Ostoin, 2007 for details). Participants were presented with a reference map and a camera view map and asked to associate terrain features presented in the reference map to identify the direction of travel in the camera view map. Participants were presented with 24 practice trials with feedback and then 24 experimental trials. The practice trials increased in difficulty, as the first 8 trials presented 4 response options, the second 8 trials presented 8 response options, and the last 8 trials presented 12 response options. All experimental trials presented 12 response options. The dependent variable was the root mean squared error (RMSE) of the reported direction compared to the correct direction of travel.

8.1.3 | Raven advanced progressive matrices

See Study 1.

8.1.4 | Antisaccade

See Study 1.

8.1.5 | Visual arrays

See Study 1.

8.1.6 | Mental counters

See Study 2.

8.2 | Data analysis

Data analysis procedures were the same as in Studies 1 and 2.

9 | RESULTS AND DISCUSSION

A summary of the cognitive ability measures is listed in Table 2. The measures were rather normally distributed with low values for kurtosis and skewness. There was also a rather broad range of ability in each measure. The

TABLE 10 Correlations among measures in Study 3

Measure	1	2	3	4	5	6	7	8	9	10
1. Pupil—black M	–									
2. Pupil—gray M	**0.74**	–								
3. Pupil – white	**0.57**	**0.76**	–							
4. Pupil—black SD	**−0.19**	−0.05	0.03	–						
5. Pupil—gray SD	**0.15**	**0.18**	**0.10**	**0.42**	–					
6. Pupil—white SD	0.07	**0.13**	**0.15**	**0.35**	**0.48**	–				
7. Antisaccade	−0.05	−0.06	−0.03	−0.02	−0.10	−0.09	–			
8. Visual arrays	0.00	0.04	0.04	−0.05	−0.02	−0.02	**0.41**	–		
9. Terrain orientation	−.10	−0.03	0.00	**0.13**	**0.10**	**0.14**	**−0.39**	**−0.38**	–	
10. Mental counters	0.06	0.04	0.01	−0.08	**−0.10**	**−0.12**	**0.43**	**0.45**	**−0.46**	–
11. Raven	0.04	0.03	0.08	−0.08	−0.04	−0.07	**0.38**	**0.43**	**−0.47**	**0.52**

Note: Bolded correlations are significant $p < .05$.

Abbreviations: M, mean pupil diameter; SD, standard deviation of pupil diameter (hippus).

cognitive measures were all significantly intercorrelated (see Table 10).

Figure 1 shows the distributions of pupil diameter against each background and Figure 2 shows the distributions for pupillary hippus. Descriptive statistics for the pupil measures are listed in Table 3. Pupil diameters were significantly smaller against a white background screen ($b = -0.68$, 95% CI = [−0.73, −0.62], $p < .001$) and significantly larger against a black background screen ($b = 1.50$, 95% CI = [1.45, 1.56], $p < .001$) compared to a gray background screen.

There was also more interindividual variability across participants with the black background screen. Specifically, there was significantly more interindividual variability in pupil diameter against the black screen compared to the gray screen ($t[438] = 20.17$, $p < .001$), and significantly more interindividual variability with the gray screen compared to the white screen ($t[440] = -8.74$, $p < .001$). This replicates Studies 1 and 2 and Tsukahara and Engle (2021). The black background also produced significantly greater pupillary hippus compared to the gray screen ($b = 0.06$, 95% CI = [0.04, 0.08], $p < .001$), and the white screen produced significantly less pupillary hippus than the gray screen ($b = -0.03$, 95% CI = [−0.05, −0.02], $p < .001$). There was also more interindividual variability in pupillary hippus across the different background screens. Specifically, there was more variability in pupillary hippus against the black screen compared to the gray screen ($t[426] = 13.61$, $p < .001$), and there was more variability in pupillary hippus against the gray screen compared to the white screen ($t[430] = -6.65$, $p < .001$). Collectively, these results all confirm the idea that dark luminance conditions maximize interindividual variability.

We next examined correlations between the pupil measures and cognitive ability measures. Overall, the correlations were mostly small and most were not significant. There was only one significant correlation between pupil diameter and the cognitive measures: a significant negative correlation between the terrain orientation task and pupil diameter against the white background. The terrain orientation task was also positively correlated with pupillary hippus against all three backgrounds. Note, for the terrain orientation task, lower values indicate better performance (less error). Finally, there were significant negative correlations between pupillary hippus against the white and gray backgrounds and performance on the mental counters task. But all the significant correlations were also low in magnitude.

As the next step, we examined whether the relations between pupil measures and cognitive ability measures were moderated by background screen color. Summaries of the models on pupil diameter are listed in Table 11. Only the terrain orientation task showed a significant interaction with background color, as the black background had a stronger correlation with pupil diameter. No other main effects or interactions were significant. Summaries of the models on pupillary hippus are listed in Table 12. There was a significant main effect of terrain orientation performance, but no other main effects or interactions were significant.

Finally, we performed the same factor-level analysis as we did in Study 2. All the cognitive measures were allowed to load onto a factor, and pupil diameter and pupillary hippus from each background screen were allowed to load onto a factor (see Figure 4). The model fit the data well ($\chi^2[41] = 154.20$, CFI = 0.93, TLI = 0.91, RMSEA = 0.08, SRMR = 0.05). In this case, the correlation between

TABLE 11 Summaries of linear mixed effect models on pupil diameter in Study 3

Effect	*B*	*SE*	*p*
Antisaccade	−0.05	0.05	.34
Antisaccade × background	0.01	0.02	.64
Visual arrays	0.03	0.05	.60
Visual arrays × background	0.02	0.02	.28
TOT	−0.05	0.04	.29
TOT × background	**0.05**	**0.02**	**.007**
Mental counters	0.03	0.04	.41
Mental counters × background	−0.03	0.02	.14
Raven	0.04	0.04	.31
Raven × background	0.02	0.02	.36

Abbreviations: *B*, standardized regression coefficient; *SE*, standard error of estimate; TOT, terrain orientation task. Bolded values are significant $p < .05$.

TABLE 12 Summaries of linear mixed effect models on pupillary hippus in Study 3

Effect	*B*	*SE*	*p*
Antisaccade	−0.01	0.01	.17
Antisaccade × background	0.00	0.01	.86
Visual arrays	−0.01	0.01	.34
Visual arrays × background	0.00	0.01	.35
TOT	**0.02**	**0.01**	**.001**
TOT × background	−0.01	0.00	.08
Mental counters	**−0.01**	**0.01**	**.01**
Mental counters × background	0.00	0.00	.50
Raven	−0.01	0.01	.09
Raven × background	0.01	0.00	.20

Abbreviations: *B*, standardized regression coefficient; *SE*, standard error of estimate; TOT, terrain orientation task.
Bolded values are significant at $p < .05$.

cognitive ability and pupillary hippus was significantly negative, albeit small. There was no significant correlation between pupil diameter and cognitive ability. For the individual measures, the cognitive ability factor did not significantly correlate with pupil diameter against the black ($r[445] = .08$, $p = .11$), gray ($r[445] = .04$, $p = .37$), or white background ($r[445] = .05$, $p = .32$). Correlations with pupillary hippus were small, and only significant for the white background (black: $r(445) = -.07$, $p = .14$, gray: $r(445) = -.09$, $p = .05$, white: $r(445) = -.10$, $p = .04$).

9.1 | Combined analysis

Because the studies shared several tasks, we combined the three datasets to specify a factor-analytic model of the association between the pupil measures and cognitive abilities. For the cognitive ability factor, we allowed any measure that was included in more than one study to load onto the factor. When specifying the model, we allowed all available pairwise correlations to inform the sample variance-covariance matrix, thus maximizing our sample size. The resulting model is depicted in Figure 5. The model fit the data acceptably ($\chi^2[52] = 276.94$, CFI = 0.91, TLI = 0.89, RMSEA = 0.07, SRMR = 0.06). There was not a significant correlation between pupil diameter and cognitive ability factors, but there was a small, significant negative correlation between pupillary hippus and cognitive ability.

9.2 | Potential confounds

9.2.1 | Age

Pupil diameter systematically declines with age in adulthood (Bak et al., 2017; Birren et al., 1950; Ko et al., 2011; Winn et al., 1994). The present sample comprised mostly young adults, and it had similar characteristics to university samples. But to rule out the potential that age was masking our results, we estimated correlations between age and pupil diameter in the present studies. Age did not correlate with pupil diameter or pupillary hippus (all $|r|\text{s} < .08$). Therefore, we do not believe age was confounded with pupil size, given the narrow age range in the data.

9.2.2 | Data missingness

Another potential confound was missingness. It is possible that participants with more missing data had higher values for pupillary hippus, and that may have driven any association between cognitive ability and pupillary hippus. Data can be missing in eye-tracking for any number of reasons, including blinks, eye tracker malfunction, offscreen fixations, or a participant's head moving outside the trackable range. After removing participants who had more than 40% missing data, we examined correlations between cognitive ability, pupil diameter, pupillary hippus, and missingness by simply adding the proportion of missing data as a manifest variable to the model in Figure 5. There was a significant negative correlation between missingness and pupil diameter ($r = -.27$, $p < .001$), a significant positive correlation between missingness and pupillary hippus ($r = .44$, $p < .001$), and a significant negative correlation between cognitive ability and missingness ($r = -.13$, $p = .002$). However, when entered into a structural regression model predicting cognitive ability, only

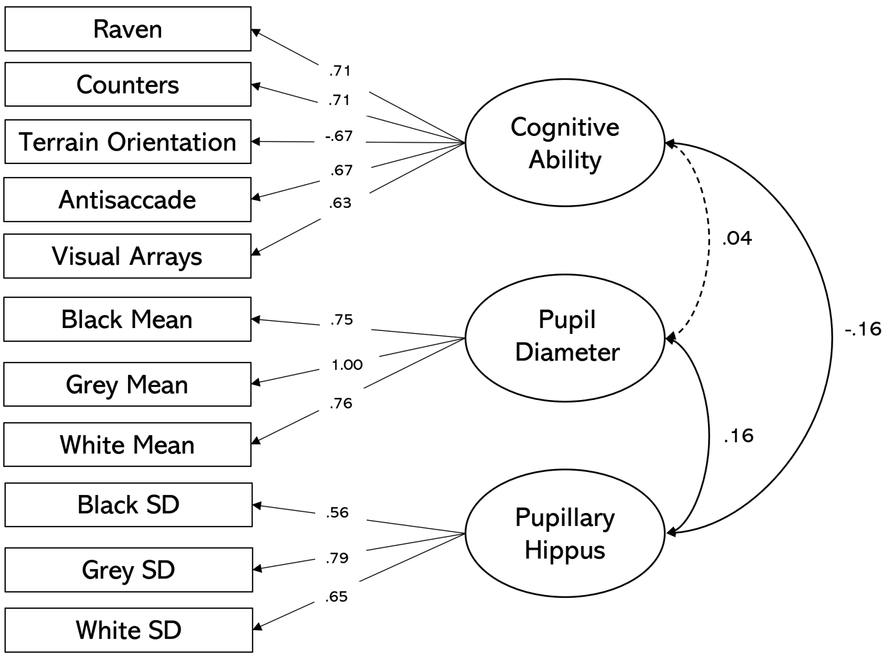

FIGURE 4 Latent factor analysis of the relations among Cognitive Ability, Pupil Diameter, and Pupillary Hippus in Study 3. Solid lines indicate significant paths at $p < .05$, dashed lines indicate non-significant paths

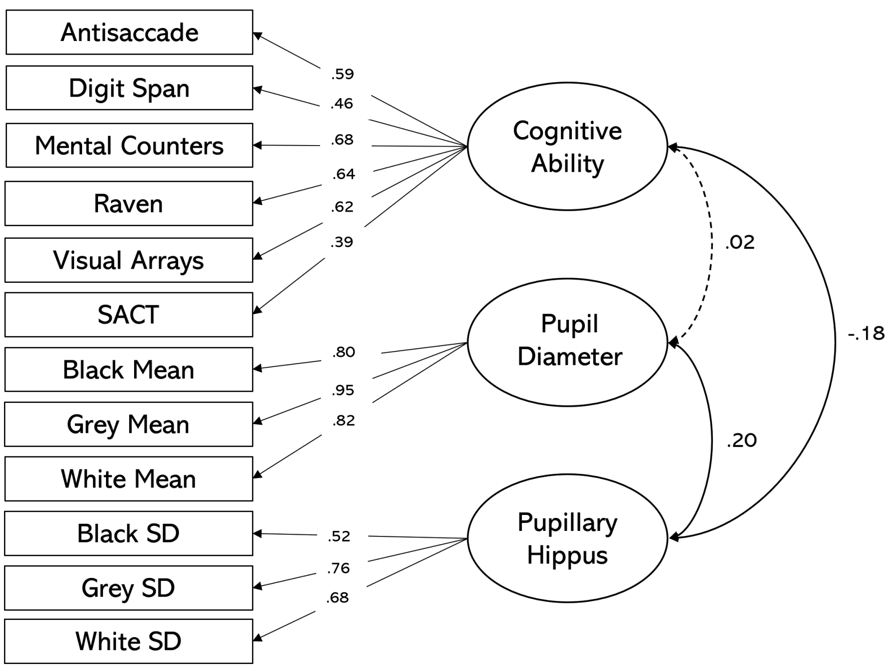

FIGURE 5 Latent factor analysis of the relations among Cognitive Ability, Pupil Diameter, and Pupillary Hippus in combined data. Solid lines indicate significant paths at $p < .05$, dashed lines indicate non-significant paths

pupillary hippus had a significant direct effect. That is, even after controlling for the amount of missing data participants had, there was still a small but significant negative association between pupillary hippus and cognitive ability.

9.2.3 | Racial/ethnic differences in pupil size

Previously, racial/ethnic differences in pupil diameter have been shown to affect the correlation between pupil size and cognitive ability (Tsukahara & Engle, 2021; Unsworth et al., 2021). To examine the effect of race/ethnicity on pupil size, we compared mean pupil diameter across race/ethnicities, combining data across Studies 2 and 3 (no ethnicity data were collected in Study 1). These data should be interpreted cautiously because the samples were majority (58%) White. An ANOVA compared the Black, Hispanic, Multiracial, and White participants. Only 2 participants reported being Native American, only 22 reported being Asian, and only 8 reported being Native Hawaiian/Pacific Islander, so we only compared the four largest groups (453 White participants, 110 Multiracial participants, 94 Hispanic participants, and 78

Black participants) on mean pupil size, averaged across background screens. The ANOVA indicated a small but significant main effect of race/ethnicity (F[3, 667] = 3.35, p = .02). Follow-up Bonferroni-corrected t tests indicated that participants who identified as Hispanic had the largest resting pupil diameters (M = 3.49 mm), which were significantly larger than the Multiracial participants (M = 3.26 mm, p = .01). There was not a significant difference between Hispanic and Black (M = 3.31), Hispanic and White (M = 3.35) participants, nor between any other groups (all ps > .10). Hispanic participants (M = 1.08, SD = 0.27) also had significantly higher pupillary hippus, overall, compared to Black participants (M = 0.93, SD = 0.27) and White participants (M = 0.94, SD = 0.28). There was not a significant difference between Hispanic participants and multiracial participants (M = 0.96, SD = 0.30, p = .06) Finally, residualizing pupil diameter and pupillary hippus (SD) on race and estimating the latent model yields similar results, with a small negative correlation between cognitive ability and hippus (r = −.15, p = .01) and a near-zero correlation between cognitive ability and mean pupil size (r = .03, p = .67). Therefore, the correlations among cognitive ability, mean pupil size, and pupillary hippus seemed unaffected by racial/ethnic differences in the present study.

9.2.4 | Education level

Next, we examined whether there were any confounding effects of education level. A z score composite of cognitive ability (antisaccade, visual arrays, terrain orientation, mental counters, and visual arrays) did not differ between the three best-represented education levels (309 college graduates, 132 participants with some college, and 297 participants with a high school diploma), F(2, 735) = 2.64, p = .07. Mean pupil size also did not differ across education levels, F(2, 672) = .28, p = .75), but pupillary hippus did, F(2, 672) = 11.41, p < .001. Bonferroni-corrected follow-up t tests indicated that participants with a high school diploma (M = 1.02, SD = 0.30) had significantly higher pupillary hippus than participants with a college degree (M = 0.91, SD = 0.28, p < .001). There was no difference between participants with a high school diploma and some college (M = 0.96, SD = 0.28, p = .12) nor between participants with some college and a college degree (p = .33). Finally, estimating the latent variable model separately for each group indicated that among participants with a high school education only, there was neither a significant correlation between cognitive ability and pupil size (r = .02, p = .82), nor between cognitive ability and pupillary hippus (r = −.02, p = .81). In the participants with some college education, there was not a significant correlation between pupil size and cognitive ability (r = .08, p = .67), but there was a significant negative correlation between cognitive ability and pupillary hippus (r = −.32, p = .04). Finally, in participants with a college degree there was neither a significant correlation between pupil size and cognitive ability (r = .04, p = .61) nor between pupillary hippus and cognitive ability (r = −.15, p = .13). Thus, the correlation between hippus and ability was only present in participants with some college education, which was the smallest of the three groups represented in the analysis, and in no individual group was there a correlation between pupil size and cognitive ability.[7]

9.2.5 | Fixation instability

One of the factors that could be driving intraindividual variability in pupil diameter (i.e., hippus), could be the degree to which people move their eyes during the baseline measure. To examine this, we computed a measure of fixation instability by taking the standard deviation of gaze position in the horizontal direction and the standard deviation of gaze position in the vertical direction and averaging these values for each participant (Di Russo et al., 2003; Unsworth et al., 2019). Higher values for this metric indicate that a person's gaze moves around the screen more often. We added this measure as a covariate in the latent analysis of cognitive ability, mean pupil diameter, and pupillary hippus in the combined sample. Indeed, fixation instability correlated positively with pupillary hippus (r = .45, p < .001). Fixation instability also negatively correlated with cognitive ability, but the correlation was very small (r = −.11, p = .01). Finally, fixation instability did not mediate the effect of pupillary hippus on cognitive ability in multiple regression. Therefore, although fixation instability did seem to lead to higher values of intraindividual pupil variability, it did not appear to be a significant confound.

10 | GENERAL DISCUSSION

The goal of the present study was to examine correlations between baseline pupil measures and cognitive ability. In each of three separate studies, we tested a large sample of U.S. military service members on a battery of cognitive tasks in addition to measuring resting/baseline pupil size and resting pupillary hippus. Based on recent work showing that interindividual variability

[7]We would like to thank an anonymous reviewer for recommending this analysis.

in pupil size and correlations with cognitive measures can be moderated by factors such as screen background color and luminance, we measured pupil size as participants fixated on a screen that went from black to gray to white in 30-s intervals. Replicating what Tsukahara and Engle (2021) observed, there was a significant increase in interindividual variability in pupil size and pupillary hippus for black screens compared to white and gray screens. Thus, the black background screen created the most diversity in the sample in both pupillary measures of interest.

Overall, we observed little evidence for a correlation between resting pupil size and cognitive ability. At both the task and factor levels, correlations between pupil size and cognitive abilities were largely weak and non-significant, even against the black background, which produced the most variability across participants. There was also only mixed evidence moderation of the correlations between cognitive measures and pupil size by background screen color. Specifically, there was significant moderation of the relation between measures of pupil size and cognitive ability for some measures in some studies, but there was not a clear pattern indicating that pupil measures taken against a black background produced stronger correlations with cognitive abilities than pupil measures taken against a white or gray background.

In addition to examining pupil diameter, we also measured pupillary hippus—the degree to which one's pupil spontaneously fluctuates across a window of time. We computed this measure separately for each background screen. Interestingly, this measure negatively correlated with cognitive ability at the factor level in Studies 1 and 3 and the combined dataset, although the correlation was consistently small and was not significant in Study 2. Further, when controlling for racial/ethnic differences and separately examining the correlation for participants at different levels of education, the correlation was not consistently observed. Therefore, we do not believe that resting pupillary hippus is an individual difference that reliably correlates with cognitive abilities. Rather, we believe pupillary dynamics within and during attention and memory tasks carry important information. Specifically, intraindividual variability in pupil diameter across trials has been shown to negatively correlate with working memory capacity (Robison & Brewer, 2020; Unsworth & Robison, 2015, 2017a), attention control (Unsworth & Robison, 2017a; Robison & Brewer, 2022), long-term memory abilities (Madore et al., 2020; Robison et al., 2022), and positively correlate with self-reported instances of mind-wandering and distraction (Unsworth & Robison, 2017a; Robison & Brewer, 2020, 2022), and self-reported media multitasking (Madore et al., 2020).

In addition to tonic arousal regulation, pupillary measures of phasic responsiveness tend to positively correlate with working memory capacity (Unsworth & Robison, 2015, 2017a), attention control (Unsworth & Robison, 2017a), and long-term memory (Robison et al., 2022). Unsworth and Robison (2017b) propose that some people experience more fluctuations in arousal because of a structural or functional difference in the connection between the LC and large-scale cortical networks that implement externally focused goal-directed cognition (e.g., frontoparietal control network, salience network) and those that support internally directed cognition (e.g., default-mode network). Further, in moments when attention needs to be allocated toward encoding information into long-term memory, storing it in working memory, or emitting a fast and accurate response, phasic responsiveness is required. People who do this with relatively more effectiveness tend to have higher cognitive ability.

Although the present data suggest there may also be a small negative relation between resting pupillary hippus and cognitive abilities, the relation was not consistently observed in groups in the present samples. Further, this same has not been observed in other studies (Robison & Brewer, 2022; Tsukahara & Engle, 2021; Unsworth et al., 2019). Therefore, any correlation between resting pupillary hippus and cognitive ability is small, if present at all. Thus, we would argue that the in-task pupillary dynamics, which can be used to index tonic arousal regulation and phasic responsiveness, are more useful from an individual-differences perspective than measures of pupillary dynamics at rest.

10.1 | Potential limitations

The present study had several strengths including multiple studies with several reliable cognitive and physiological measures drawn from the large and diverse sample. However, there are several potential limitations worth mentioning. First, because the sessions were not designed to later estimate working memory capacity, fluid intelligence, or attention control at the latent level, the reported set of tasks was limited to a smaller set of tasks per study, and we were unable to specify factors for individual cognitive abilities like working memory capacity, fluid intelligence, or attention control. Generally, it is advantageous to examine cognitive abilities with multiple measures of each construct, then capture shared variance in performance across tasks using factor analyses. In the present studies, we were only able to specify a general cognitive ability factor that represented shared variance among all the cognitive measures. However, the Raven, antisaccade, SACT, and visual arrays tasks have been used previously in latent-level analyses of fluid intelligence, attention control, and working/short-term memory (Draheim et al., 2021; Kane et al., 2016; Shipstead et al., 2014, 2015; Unsworth et al., 2014, 2020; Unsworth & McMillan, 2014; Unsworth & Spillers, 2010), and they are standard measures of their respective constructs. Further, when allowed to load onto a general Cognitive Ability factor, the tasks all significantly loaded onto that factor, and thus the sample covariance structure was well-captured by the model. Therefore, we believe this general factor did well to capture a broad cognitive ability, similar to the "Common" factor specified by Tsukahara and Engle (2021) in their recent study.

Another potential limitation of the current sample was our convenience sampling method. Our sample came from a specific population: U.S. military officers and enlisted service members on a military base, most of whom were young adults. Active duty U.S. military membership is largely male (83%) and ethnically white (69%; U.S. Department of Defense, 2019). Our sample matched this gender composition and was slightly more ethnically diverse than active-duty U.S. military overall. Regardless, there are concerns about convenience samples and the generalizability of findings from them. Specifically, convenience samples can lead to range restriction if the sample is from a specific segment of the population (e.g., largely high-ability, ethnically homogenous, etc.). For the purposes of the present study, the two major concerns were whether we had a range restriction of cognitive ability and/or range restriction of pupil size. Based on comparisons of distributions, our sample was as diverse, and in some cases more diverse in cognitive ability, than the samples from Tsukahara and Engle (2021), which included students from two universities and community participants (see Supplemental Materials). Thus, we do not believe range restriction precluded us from observing relations between pupil size and cognitive ability. Indeed, we specifically designed Study 3 to include as wide a range of cognitive ability as is feasible in a military setting by sampling from several different occupations that require very different cognitive-testing scores to qualify for their positions.

A third potential limitation is our use of newer low-cost eye trackers compared to other models. The eye-trackers used in the present study cost about USD$2000, as opposed to some models that can run in the USD$20,000–30,000 range. Therefore, there could be a concern that the less-expensive models collect noisier data. We do not believe this to be the case for several reasons. First, the measures (mean and hippus) showed high internal consistencies, with split-half corrected coefficients at or above .99. Second, recent research has specifically examined the fidelity of data from Gazepoint GP3 and GP3HD eye-trackers and found that they produce as precise and reliable estimates of pupil size and gaze position as more expensive models (Brand, Diamond, Thomas, & Gilbert-Diamond, 2022; Cuve, Stojanov, Roberts-Gaal, Catmur, & Bird, 2022; Mannaru et al., 2017). Third, all participants wore a pair of plastic eyeglass frames (without lenses) with a fiducial marker of known size taped to the center of the frames. The eye tracker automatically converted pupil size measurements to millimeters based on a pixel-to-millimeter conversion with reference to the fiducial marker's size in pixels at each sample. Therefore, we felt certain that the data coming from the GP3HD eye-trackers used in the present study were providing valid and reliable indices of pupil diameter.

10.2 | Conclusions

We did not observe evidence for a relation between resting pupil size and cognitive ability, nor consistent evidence for a moderation of the relation between pupil size and cognitive ability by fixation screen background color. We replicated prior work by Tsukahara and Engle (2021) showing that interindividual variability in pupil size does indeed increase with darker environmental luminance conditions. But even under these conditions, correlations between pupil size and cognitive abilities were largely small and in most cases statistically indistinguishable from zero. We had a large sample size with a broad range of both pupil size and cognitive ability, and we collected multiple cognitive measures tapping into attention control, working memory capacity, visual short-term memory capacity, fluid intelligence, and visuospatial abilities. At the factor level, the correlation between general cognitive ability and pupil size was small and not statistically significant, even with our large sample size. Therefore, we would urge future research to rigorously replicate and demonstrate a consistent correlation between cognitive ability and resting pupil size, especially the relation between intelligence and pupil size, before making any strong theoretical claims about such a relation.

AUTHOR CONTRIBUTIONS

Joseph Coyne: Conceptualization; methodology; project administration; supervision; writing – review and editing. **Ciara Sibley:** Conceptualization; investigation; methodology; project administration; supervision; writing – review and editing. **Noelle L Brown:** Conceptualization; methodology; project administration; supervision; writing – review and editing. **Brittany Neilson:** Data curation; project administration; writing – review and editing. **Cyrus K Foroughi:** Conceptualization; methodology; project administration; writing – review and editing.

ACKNOWLEDGMENTS

We would like to thank the research personnel and participants who contributed to the data collection portion of this project.

FUNDING INFORMATION

The research team did not receive any specific grant funding for this project.

CONFLICT OF INTEREST

None.

ORCID

Matthew K. Robison https://orcid.org/0000-0003-3565-6002

REFERENCES

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology*, *9*(1), 5–37. https://doi.org/10.1207/s15327876mp0901_1

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*(63), 63. https://doi.org/10.12688/wellcomeopenres.15191.1

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

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*. https://github.com/crsh/papaja

Bak, E., Yoo, Y. J., Yang, H. K., & Hwang, J. M. (2017). Quantitative pupillometry of the pupillary light reflex in Koreans. *Journal of the Korean Ophthalmological Society*, *58*(6), 712–717. https://doi.org/10.3341/jkos.2017.58.6.712

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Birren, J. E., Casperson, R. C., & Botwinick, J. (1950). Age changes in pupil size. *Journal of Gerontology*, *5*(3), 216–221. https://doi.org/10.1093/geronj/5.3.216

Brand, J., Diamond, S. G., Thomas, N., & Gilbert-Diamond, D. (2022). Evaluating the data quality of the Gazepoint GP3 low-cost eye tracker when used independently by study participants. *Behavior Research Methods*, *53*(4), 1502–1514. https://doi.org/10.3758/s13428-020-01504-2

Champely, S. (2018). *PairedData: Paired data analysis*. https://CRAN.R-project.org/package=PairedData

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences*, *106*(14), 5529–5533. https://doi.org/10.1073/pnas.0810002106

Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2022). Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods*, *54*(2), 1027–1049. https://doi.org/10.3758/s13428-021-01654-x

Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2022). Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods*, *54*(2), 1027–1049 https://doi.org/10.3758/s13428-021-01654-x

Di Russo, F., Pitzalis, S., & Spinelli, D. (2003). Fixation stability and saccadic latency in elite shooters. *Vision Research*, *43*(17), 1837–1845. https://doi.org/10.1016/S0042-6989(03)00299-2

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242–275. https://doi.org/10.1037/xge0000783

Elsmore, T. F. (1994). SYNWORK1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 421–426. https://doi.org/10.3758/BF03204659

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167. https://doi.org/10.1002/acp.1624

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*, 119–129. https://doi.org/10.1111/j.1469-8986.2007.00605.x

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

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 645–662. https://doi.org/10.1037/0278-7393.33.4.645

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

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

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of*

Experimental Psychology: General, *145*(8), 1017–1048. https://doi.org/10.1037/xge0000184

Kawasaki, A., Ouanes, S., Crippa, S. V., & Popp, J. (2020). Early-stage Alzheimer's disease does not alter pupil responses to colored light stimuli. *Journal of Alzheimer's Disease*, *75*(4), 1273–1282. https://doi.org/10.3233/JAD-200120

Ko, B. U., Ryu, W. Y., & Park, W. C. (2011). Pupil size in the normal Korean population according to age and illuminance. *Journal of the Korean Ophthalmological Society*, *52*(4), 401–406. https://doi.org/10.3341/JKOS.2011.52.4.401

Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality*, *80*, 17–22. https://doi.org/10.1016/j.jrp.2019.03.007

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Learn, J. R. (2021). Pupil size may be linked to intelligence. *Discover*.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. https://doi.org/10.1038/36846

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*(7832), 87–91. https://doi.org/10.1038/s41586-020-2870-z

Mäki-Marttunen, V., Hagen, T., Aminihajibashi, S., Foldal, M., Stavrinou, M., Halvorsen, J. H., Laeng, B., & Espeseth, T. (2018). Ocular signatures of proactive versus reactive cognitive control in young adults. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(5), 1049–1063. https://doi.org/10.3758/s13415-018-0621-5

Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., & Coyne, J. T. (2017). Performance evaluation of the Gazepoint GP3 eye tracking device based on pupil dilation. In *International Conference on Augmented Cognition* (pp. 166–175). Springer.

Mather, M., & Harley, C. W. (2016). The locus coeruleus: Essential for maintaining cognitive function and the aging brain. *Trends in Cognitive Sciences*, *20*(3), 214–226. https://doi.org/10.1016/j.tics.2016.01.001

McLaren, J. W., Erie, J. C., & Brubaker, R. F. (1992). Computerized analysis of pupillograms in studies of alertness. *Investigative Ophthalmology & Visual Science*, *33*(3), 671–676.

Ostoin, S. D. (2007). *An Assessment of the Performance-based Measurement Battery (PBMB), the Navy's Psychomotor Supplement to the Aviation Selection Test Battery (ASTB)*. Naval Postgraduate School Monterey CA.

Rajkowski, J. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of lc in attention. *Society for Neuroscience Abstracts*, *19*(1–3), 974.

Revelle, W. (2018). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University https://CRAN.R-project.org/package=psych

Robison, M. K., & Brewer, G. A. (2022). Individual differences in working memory capacity, attention control, fluid intelligence, and pupillary measures of arousal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://psycnet.apa.org/doi/10.1037/xlm0001125

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., Trost, J. M., Schor, D., Gibson, B. S., & Healey, M. K. (2022). Pupillary correlates of individual differences in long-term memory. *Psychonomic Bulletin & Review*, 1–12 https://doi.org/10.3758/s13423-022-02081-5

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36 http://www.jstatsoft.org/v48/i02/

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

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, *77*, 1863–1880. https://doi.org/10.3758/s13414-015-0899-0

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Advances in psychology* (Vol. *65*, pp. 31–86). Elsevier. https://doi.org/10.1016/S0166-4115(08)61218-2

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tsukahara, J. S., Burgoyne, A. P., & Engle, R. W. (2021). Pupil size is a marker of intelligence. *Scientific American* https://www.scientificamerican.com/article/pupil-size-is-a-marker-of-intelligence/

Tsukahara, J. S., & Engle, R. W. (2021). Is baseline pupil size related to cognitive ability? Yes (under proper lighting conditions). *Cognition*, *211*, 104643. https://doi.org/10.1016/j.cognition.2021.104643

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

U.S. Department of Defense. (2019). Demographics report. *Military OneSource* https://download.militaryonesource.mil/12038/MOS/Reports/2019-demographics-report.pdf

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, *150*, 14–25. https://doi.org/10.1016/j.actpsy.2014.04.001

Unsworth, N., Miller, A. L., & Robison, M. K. (2021a). Is working memory capacity related to baseline pupil diameter? *Psychonomic Bulletin & Review*, *28*(1), 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*(3), 757–765. https://doi.org/10.3758/s13423-014-0747-6

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/xlmsss0000421

Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective, & Behavioral Neuroscience*, *19*, 1074–1093. https://doi.org/10.3758/s13415-019-00709-z

Unsworth, N., Miller, A. L., & Robison, M. K. (2021). No consistent correlation between baseline pupil diameter and cognitive abilities after controlling for confounds—A comment on. *Cognition*, *215*, 104825.

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

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. https://doi.org/10.1523/JNEUROSCI.0454-15.2015

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503. https://doi.org/10.1038/nature04171

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York http://ggplot2.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pederson, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. https://CRAN.R-project.org/package=cowplost

Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, *35*(3), 1132–1137.

How to cite this article: Robison, M. K., Coyne, J. T., Sibley, C., Brown, N. L., Neilson, B., & Foroughi, C. (2022). An examination of relations between baseline pupil measures and cognitive abilities. *Psychophysiology*, *00*, e14124. https://doi.org/10.1111/psyp.14124