Cognition 211 (2021) 104643

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Is baseline pupil size related to cognitive ability? Yes (under proper lighting conditions)

Jason S. Tsukahara *, Randall W. Engle

Georgia Institute of Technology, United States of America

ARTICLE INFO

Keywords:
Fluid intelligence
Working memory capacity
Pupil size
Luminance

ABSTRACT

There has been some controversy as to whether baseline pupil size is related to individual differences in cognitive ability. Previously, we had shown that a larger baseline pupil size was associated with higher cognitive ability and that the correlation to fluid intelligence was larger than that to working memory capacity (Tsukahara, Harrison, & Engle, 2016). However, other researchers have not been able to replicate our findings – though they only measured working memory capacity and not fluid intelligence. Many of the studies showing no relationship had major methodological issues, namely small baseline pupil size values – down to the physiological minimum – that resulted in reduced variability on baseline pupil size. We conducted two large-scale studies to investigate how different lighting conditions affect baseline pupil size values and the correlation with cognitive abilities. We found that fluid intelligence, working memory capacity, and attention control did correlate with baseline pupil size except in the brightest lighting conditions. We showed that a reduced variability in baseline pupil size values is due to the monitor settings being too bright. Overall, our findings demonstrated that the baseline pupil size – working memory capacity relationship was not as strong or robust as that with fluid intelligence or attention control. Our findings have strong methodological implications for researchers investigating individual differences in task-free or task-evoked pupil size. We conclude that fluid intelligence does correlate with baseline pupil size and that this is related to the functional organization of the resting-state brain through the locus coeruleus-norepinephrine system.

1. Introduction

How the pupil changes in size with mental effort and various cognitive processes has been a prolific area of research (Ahern & Beatty, 1979; Beatty, 1982; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Heitz, Schrock, Payne, & Engle, 2008; Hess & Polt, 1964; Kahneman & Beatty, 1966; Ullwer et al., 2003; van der Meer et al., 2010). Even more so, the ease of obtaining pupillary measures has provided a non-invasive method for psychologists to understand the neural underpinnings of these cognitive processes (Aston-Jones & Cohen, 2005; Joshi & Gold, 2019; Joshi, Li, Kalwani, & Gold, 2016; Laeng, Sirois, & Gredeback, 2012; Murphy, Robertson, Balsters, & O'Connll, 2011; Rajkowski, Kubiak, & Aston-Jones, 1993; Ruud Van Den Brink, Murphy, & Nieuwenhuis, 2016; Sara, 2009). More recently, researchers have shown cognitive-related changes in pupil size to be associated with individual differences in cognitive abilities, such as intelligence and working memory capacity (Tsukahara et al., 2016; Ullwer et al., 2003; Unsworth & Robison, 2015, 2016, 2017a, 2017b; Unsworth, Robison, & Miller, 2019; van der Meer et al., 2010).

However, studying individual differences in pupil size has been met with mixed findings. In particular, there has been mixed success replicating a finding we reported; that task-free baseline pupil size is related to fluid intelligence and working memory capacity (Tsukahara et al., 2016). Though the original finding was incidental (Heitz et al., 2008), we followed it up over the years and found, in three experiments, that a larger baseline pupil size was consistently related to higher working memory capacity even after controlling for effort, familiarity with the environment, age, and other confounds (Tsukahara et al., 2016). Using a large sample with the full range of abilities we found that fluid intelligence, not working memory capacity, uniquely predicted baseline pupil size (Experiment 3 of Tsukahara et al., 2016). After repeatedly finding the same result in our lab over the years we were confident that this effect was real and robust to various confounds such as age. We were also not alone in finding this relationship between baseline pupil size and intelligence (Bornemann, Foth, & Horn, 2010; Ullwer et al., 2003; van der Meer et al., 2010).

* Corresponding author at: School of Psychology, Georgia Institute of Technology, 654 Cherry St., Atlanta, GA 30332, United States of America.
E-mail address: jason.tsukahara@gatech.edu (J.S. Tsukahara).

https://doi.org/10.1016/j.cognition.2021.104643
Received 26 March 2020; Received in revised form 9 February 2021; Accepted 19 February 2021

Based on our finding, and a body of research linking pupil size to specific brain regions (Aston-Jones & Cohen, 2005), we proposed that individual differences in fluid intelligence is related to the locus-coeruleus norepinephrine system in the brain (Tsukahara et al., 2016). Specifically, that fluid intelligence is related to the functional organization of the resting-state brain arising from the neuromodulatory role of the locus coeruleus-norepinephrine system. Other researchers have proposed similar yet distinct theories as to how higher-order cognitive abilities are related to the locus coeruleus-norepinephrine system (Unsworth & Robison, 2017a; van der Meer et al., 2010).

However, some researchers have not been able to replicate our findings, specifically with working memory capacity (Aminihajibashi, Hagen, Foldal, Laeng, & Espeseth, 2019; Unsworth et al., 2019; Unsworth & Robison, 2015, 2017b). In a recent meta-analysis Unsworth, Miller, and Robison (2020), it was found that across ~30 studies the only ones to find a significant correlation between baseline pupil size and working memory capacity were the ones from our lab. This is concerning to say the least. Why would this finding that repeatedly replicated in our lab not replicate for other researchers? We will now consider some possible reasons.

1.1. Is baseline pupil size related to fluid intelligence?

In Tsukahara et al. (2016) we emphasized that it is fluid intelligence, not working memory capacity that is related to baseline pupil size. Although working memory capacity and fluid intelligence are highly correlated at the latent level, $r \sim 0.6$–0.8, they are considered to be distinct abilities (Conway, Kane, & Engle, 2003; Kane, Hambrick, & Conway, 2005). Because working memory capacity and fluid intelligence are so highly correlated, they will often times show similar patterns of predictive validity to other constructs and measures, such as baseline pupil size. It is then, all too easy to equate working memory capacity with fluid intelligence. Therefore, to state that baseline pupil size is related to fluid intelligence (Tsukahara et al., 2016) should not be equated with the statement that baseline pupil size is related to working memory capacity. In fact, in Tsukahara et al. (2016) we found that the correlation of baseline pupil size with fluid intelligence ($r = 0.35$) was stronger, $p < .05$, than that with working memory capacity ($r = 0.24$). In addition, after removing their shared variance, only fluid intelligence predicted baseline pupil size.

Despite the fact that we made the claim that baseline pupil size is related to fluid intelligence we know of only a few studies that have looked at this relationship. The majority of studies have tested the baseline pupil size – working memory capacity relationship. Two studies already discussed were published before Tsukahara et al. (2016), had small sample sizes, restriction of range on ability, and used extreme groups design (Bornemann et al., 2010; van der Meer et al., 2010); yet they did find a significant relationship between baseline pupil size and performance on the Raven's advanced progressive matrices (a single measure of fluid intelligence). Although we do believe there is a stronger and more robust relationship between baseline pupil size and fluid intelligence (compared to working memory capacity), there needs to be a clearer distinction between these two highly related constructs and how they are differentially related to both task-free pupil size and task-evoked changes in pupil size. Nevertheless, there is a lack of studies on the baseline pupil size – fluid intelligence relationship despite our emphasis of fluid intelligence over working memory capacity (Tsukahara et al., 2016).

1.2. Is baseline pupil size related to working memory capacity?

Even if fluid intelligence is more strongly related to differences in pupil size, a considerable number of researchers have failed to replicate the correlation with working memory capacity (Aminihajibashi et al., 2019;Unsworth et al., 2019 ; Unsworth & Robison, 2015, 2017b). We believe there are a number of factors contributing to this. First of all, the baseline pupil size – working memory capacity correlation is not large, $r = 0.24$ (Exp. 3 from Tsukahara et al., 2016), and therefore will not necessarily be robust to potential confounds, small sample size, and measurement problems. In the meta-analysis by Unsworth et al. (2020), of the ~30 studies, almost half of them were from the Unsworth lab at University of Oregon. Of the remaining studies, not from one of our two labs, most have serious measurement problems. These primarily include small sample size (less than ~100) and measuring working memory capacity with a single task. Correlations can be inaccurate and more variable with smaller samples (Schönbrodt & Perugini, 2013) and it is highly advisable to measure a construct with multiple measures and derive common variance across them (Ackerman & Hambrick, 2020; Kovacs & Conway, 2020).

To demonstrate how using a single measure can reduce the validity of the construct let us look at data from Tsukahara et al. (2016). The composite working memory capacity (operation, symmetry, and rotation span) correlation with baseline pupil size was $r = 0.24$. The operation span correlation was $r = 0.14$, the symmetry span correlation was $r = 0.16$, the rotation span correlation was $r = 0.28$. Again, these are small effects with measures of working memory capacity. The sample size required to detect a correlation of $r = 0.14$ at $p < .05$ and power = 0.8 is $n = 397$. Or at the more optimistic correlation of $r = 0.28$, $n = 97$ is required. Therefore, most studies are barely meeting sample size requirements in the most optimistic scenario.

Nevertheless, even when these basic psychometric criteria are met, as in the Unsworth lab, there are still repeated findings of a small and non-significant correlation between baseline pupil size and working memory capacity (Unsworth et al., 2019; Unsworth & Robison, 2015, 2017b). This suggests that the baseline pupil size – working memory capacity relationship is not as robust as we previously thought.

However, it may be more complicated than this. We noticed that other labs were reporting much smaller baseline pupil size values than what we reported in our previous studies. Based on this, we suspected that it is a reduced mean and inter-individual variability on baseline pupil size values that might account for the failures to replicate. Therefore, we will now discuss how reduced inter-individual variance on baseline pupil size can impact its correlation with cognitive abilities.

1.3. Reduced variance on baseline pupil size

First, let us consider the physiological limits on the size of the pupil. This lets us establish an idea of what sort of range and variability we can expect in baseline pupil size values. Researchers have studied the light-adapted pupil reflex and can give us some insight into the range of possible pupil size values. Brown, Khanani, and Xu (2004) measured dark-adapted pupil diameter under an illumination of 1 lx. They report mean pupil diameters in the range of 5.44–8.63 mm. de Groot and Gebhard (1952) measured pupil diameter across varying levels of illumination and report mean pupil diameters in the range of 2.00–7.17 mm, from their brightest to darkest levels of illumination. Additionally, researchers will often use 2.00 mm as a minimum cutoff of acceptable and realistic pupil values in their data. Therefore, it seems pupil values around 2 mm are at the minimum physiological limit and around 9 mm are at the maximum physiological limit.

To use Unsworth et al. (2019) as a case example, both the mean baseline pupil size (3.21 mm) and standard deviation around that mean (0.49), or the inter-individual variability, were small and compared to the physiological limits determined above, this is on the lower end of possible pupil values. However, it was primarily the small inter-individual variability that caught our attention. This was concerning to us because, as the variance of a variable (e.g., baseline pupil size) decreases, so does its ability to correlate with another variable (e.g., working memory capacity). As a comparison, in Tsukahara et al. (2016) the mean baseline pupil size was 5.92 mm and the standard deviation around that mean was 1.09. In Unsworth et al. (2019), three standard deviations above the mean pupil size (4.68 mm) is still smaller than the mean pupil size value from Tsukahara et al. (2016). Tsukahara et al. (2016) did not just have a larger mean baseline pupil size, but more importantly the inter-individual variability is more than double that in Unsworth et al. (2019). More variance between individuals allows for greater chance for relationships between variables to appear and these two studies clearly display very different distributions of baseline pupil values.

Now the question is, can this lack of inter-individual variability in baseline pupil size account for the null findings in Unsworth et al. (2019) and others? There is some support for this in a study by Winn, Whitaker, Elliott, and Phillips (1994) in which they investigated the relationship between pupil size and age in different lighting conditions. They showed a larger variance on baseline pupil size (~ 9 mm–3 mm) in low luminance conditions and smaller variance (~ 6 mm–2 mm) in high luminance conditions.[1] Crucially, however, the correlation between pupil size and age decreased as the luminance increased (and pupil size variability decreased). In the darkest luminance condition, age explained 55.7% of the variance in pupil size. Whereas, in the brightest luminance condition, age only explained 21.4% of the variance in pupil size (that is more than 60% reductio in variance explained). In other words, they showed a large decrease in a known correlation between pupil size and age with a decreased variance in pupil size (due to higher luminance). This pattern is very similar to the difference in findings between Tsukahara et al. (2016) and Unsworth et al. (2019).

1.4. Reanalysis of Tsukahara et al. (2016)

Given these concerns, we decided to conduct some additional analyses on our data from Experiment 3 of Tsukahara et al. (2016) in order to test whether the small inter-individual variance in Unsworth et al. (2019) can account for the different findings between our labs. Specifically, we tested whether reducing the variance in Tsukahara et al. (2016) to a similar magnitude as Unsworth et al. (2019) would eliminate the correlation between baseline pupil size and working memory capacity.

We found that when samples with small inter-individual variance on baseline pupil size (equivalent to Unsworth et al., 2019) are drawn from a "population" (original data from Experiment 3 in Tsukahara et al., 2016) with a known correlation value between baseline pupil size and working memory capacity, $r = 0.24$, it is rare (top 95th percentile) to obtain the original correlation value and more likely to obtain a smaller correlation (Fig. S3a). It is also more likely to obtain non-significant correlations between baseline pupil size and working memory capacity (Fig. S3b). Therefore, simply reducing the inter-individual variance in baseline pupil size will more likely result in a smaller and non-significant correlation value between baseline pupil size and working memory capacity. These results suggest that the small inter-individual variance in baseline pupil size in Unsworth et al. (2019) was a major contributor to their null findings regarding a relationship between pupil size and working memory capacity.

The mixed findings and overwhelming failures of replication have been a serious concern for us. Particularly given that we continue to find a relationship between baseline pupil size and cognitive ability in our lab. We have mentioned several possible reasons we suspect other labs have shown failures to replicate the baseline pupil size cognitive ability relationship. The one that has stood out and is most consistent across failures to replicate is a reduced mean and variance on baseline pupil size. Given that most of these studies have assessed the relationship with working memory capacity and that this relationship is small, $r = 0.24$ (Tsukahara et al., 2016), it is expected that a restriction of range on pupil size values could lead to small and non-significant correlations across studies.

1.5. Current study

The most obvious reason for differences in the mean and variance of baseline pupil size values is different lighting conditions. Therefore, we conducted two studies to further investigate whether lighting conditions can lead to, in our lab, a similar reduced variance on baseline pupil size as seen in other studies and smaller and non-significant correlations to cognitive abilities. Our main hypothesis was that measuring baseline pupil size in too bright of lighting conditions will restrict the range of baseline pupil size values and thereby reduce the correlations with fluid intelligence and working memory capacity.

In addition, we also tested the relationship between baseline pupil size and attention control. Although some previous studies have not found a relationship between baseline pupil size and attention control, at the latent construct level attention control is highly related to working memory capacity and fluid intelligence (Unsworth et al., 2019). Theoretically, working memory capacity is thought to reflect differences in attention control (Engle, 2002, 2018; Engle & Kane, 2004) and the strong correlation between working memory capacity and fluid intelligence is largely, if not completely, explained by attention control (Draheim, Tsukahara, Martin, Mashburn, & Engle, 2020). However, there have been serious methodological issues with tasks used to measure attention control (Draheim, Hicks, & Engle, 2016; Hedge, Powell, & Sumner, 2018; Paap & Sawi, 2016; Rouder & Haaf, 2019). In the current study, we included a set of tasks, old and new, that have recently been shown to have high reliability and validity for measuring individual differences in attention control (Draheim et al., 2020). Therefore, we expect attention control to show a similar pattern of relationships with baseline pupil size as working memory capacity and fluid intelligence.

2. Study 1

In Tsukahara et al. (2016), baseline pupil size was measured in a dark room (2 *lx*; we obtained this measurement recently and not at the time of the study) with a black background and gray fixation on the monitor. In Unsworth et al. (2019), baseline pupil size was measured in a dimly lit room (illuminance = 30 *lx*) with a gray background and black fixation (with mean fixation luminance reported at 40 cd/m^2) on the monitor. Our main purpose, in Study 1, was to obtain a similar range of pupil values as Unsworth and Robison (2016) and evaluate the correlation with fluid intelligence, working memory capacity, and attention control. To do so, we used a moderately bright room and a gray background condition to roughly match the luminance conditions in Unsworth et al. (2019). If our hypothesis is correct, then we would expect the correlation of baseline pupil size with working memory capacity and fluid intelligence to be small and non-significant in the gray background condition but only if the range of pupil size values are restricted such as in Unsworth et al. (2019). If the range of pupil size values are not restricted, then we would not expect the correlations to be small and non-significant. For this reason, we decided to include a white background condition to ensure we had sufficiently bright luminance to reduce the range of baseline pupil size values.

2.1. Method

The data analyzed in this study was part of a larger data collection sample. The following link has a summary of the larger data collection procedure and a reference list of all publications to come out of this data collection sample with information on which tasks were used for each publication: https://osf.io/yc48s/.

2.1.1. Subjects

Subjects were college and non-college adults of the Atlanta community. They were required to be native English speakers, 18–35 years

[1] These values are estimated visually from Figure 2 of Winn et al. (1994). However, this includes older adults (up to 85 years of age). The same pattern holds when restricting the age range to a maximum of 35 years of age.

of age, and had not participated in a study with our lab before. Screening on vision was not performed. The study consisted of four 2-h sessions. Subjects were compensated with an average of $35 per session or 2 h course credit for each session. The study was approved by the Georgia Institute of Technology's Institutional Review Board under Protocol H16409.

Baseline pupil measures[2] were obtained from a total of 317 subjects with 2 subjects removed due to having too much missing pupil data on both gray and white background conditions for a final sample size of 315. See Table 1 for the demographics of the 315 subjects.

2.1.2. Tasks and procedures

Testing was conducted in a group running room with a total of 5 subject stations and one research assistant that monitored subjects and administered tasks. Three out of five of the subject stations had eye-trackers. The tasks and baseline measures were conducted on a Windows computer with an LED-backlit LCD monitor and subjects wore headphones during all tasks and baseline measures. The tasks were programmed in *E*-Prime 3.0 software (Psychology Software Tools, 2016).

We measured fluid intelligence with the *Raven's Advanced Progressive Matrices*, *letter sets*, and *number series*. Working memory capacity was measured with the *advanced versions* of the *operation span*, *symmetry span*, and *rotation span* tasks. Attention control was measured with the *antisaccade*, *selective visual arrays,*[3] and *sustained attention to cue task*. See Draheim et al. (2020) and Martin et al. (2019) for the reliability and validity of the attention control measures. Baseline pupil size was measured in two different conditions, gray and white background color, for two minutes each.

Table 1
Subject demographic for study 1 ($N = 315$).

Demographic	Category	Value
Age (Years)	Mean	22.4
	SD	4.1
Gender	Male	36%
	Female	63%
	Other/self-identify	1%
Education	Some high school	2%
	High school/GED	10%
	Some college	57%
	Associates degree	5%
	Bachelor's degree	14%
	Some graduate school	3%
	Master's degree	9%
	PhD/MD/JD/DDS	< 1%
Ethnicity[a]	White	29%
	Black or African American	33%
	Asian or Pacific Islander	29%
	Hispanic or Latino	5%
	Native American	1%
	Other	3%

[a] Other includes mixed race and other.

2.1.3. Baseline pupil measures

A SensoMotoric Instruments Red250m eye-tracker was used to record binocularly at 250 Hz. Subjects were seated approximately 65–70 cm from the monitor and did not use any head immobilization device. Before performing any task, baseline pupil size measures were obtained in the group running room and occurred on the third session of the study. The room was moderately bright (illuminance = 60 lx). An eye-tracking calibration procedure was first conducted followed by four minutes of baseline. For half of the subjects, the first two minutes had a gray background with a silver fixation and the last two minutes had a white background with a silver fixation. The order was reversed for the other half of subjects. There was a 10 s interval between the first and second baseline conditions that was excluded from analysis. Subjects were instructed that they did not have to do anything in particular during the baseline and to keep their gaze towards the monitor. Two measures were obtained from preprocessed baseline pupil data. Baseline pupil size was calculated as the average pupil size over the baseline period. Intra-individual baseline pupil variability was calculated as the standard deviation of pupil size over the baseline period (this represents variability within an individual not between individuals).

2.1.4. Fluid intelligence tasks

Raven's advanced progressive matrices (Raven, Raven, & Court, 1998). In this task subjects were presented with a matrix of figures that follow a logical pattern across rows and columns. For each problem in this task, a 3×3 matrix of 8 abstract figures was presented with the bottom-right element missing. Subjects had to identify the logical pattern and select one of eight answer choices that fits the logical pattern of the matrix. Subjects were given 10 min to solve 18 of the odd numbered problems from the full test. Scores on this task were calculated as the total number of problems solved correctly.

Letter sets (Ekstrom, French, Harman, & Dermen, 1976). Subjects were presented with 5 sets of 4-letter sequences (e.g. NOPQ DEFL ABCD HIJK UVWX). Subjects had to identify a common pattern amongst 4 of the sets and select the set of letters that did not follow the pattern (e.g. the letter sets are all in consecutive alphabetical order except for DEFL). Subjects were given 10 min to solve 30 problems. Scores on this task were calculated as the total number of problems solved correctly.

Number series (Thurstone, 1938). For each problem in this task, a series of numbers was presented that progressed in a particular logical fashion. Subjects had to identify the rule and select the next number, out of 5 answer choices, that should occur next in the series of numbers to be consistent with the logical rule. Subjects were given 5 min to complete 15 problems. Scores on this task were calculated as the total number of problems solved correctly.

2.1.5. Working memory capacity tasks

The complex span tasks consist of alternating memory storage and processing sub-tasks (Unsworth, Heitz, Schrock, & Engle, 2005). The advanced versions of the tasks included larger set-sizes of memory items (Draheim, Harrison, Embretson, & Engle, 2018). In all complex span tasks, the total number of items recalled in their correct serial position (partial score) was used to calculate scores on the task (Conway et al., 2005).

Advanced operation span. This task required subjects to remember a series of letters presented in alternation with simple math equations which they were required to solve. On each trial, subjects first solved a simple math equation followed by the presentation of a single letter. This alternation repeated until a variable set-size of letters to-be-remembered had been presented. Then, on the recall screen subjects had to recall the letters in the correct order by clicking the mouse on the appropriate letters from a matrix of letters displayed on the screen. There was a total of 14 trials (2 blocks of 7 trials), set-sizes ranged from 3

[2] The white background condition was added after the start of the data collection. Baseline pupil size was obtained from a total of 292 subjects in the white background condition.

[3] The visual arrays task is most commonly known as a visual working memory task. However, a considerable amount of behavioral and neurophysiological studies has provided evidence that performance on the selective version of this task is determined by controlled attention processes (for a review see; Martin et al., 2019). In a reanalysis of data from four different studies from our lab, we have shown that the selective version of the visual arrays strongly prefers to load onto a latent factor with attention control tasks compared to working memory capacity tasks (Martin et al., 2019). That is, even though performance on the visual arrays task is calculated with a k storage capacity score, this task shares more common variance with tasks that have minimal memory demands than tasks that place a strong demand on working memory.

to 9, and each set-size occurred twice (once in each block). Scores on the advanced operation span task were calculated using the partial-scoring method, the number of letters recalled in their correct order across all trials.

Advanced symmetry span. This task required subjects to remember a series of spatial locations in a 4 × 4 matrix presented in alternation with a pattern of squares which they had to decide whether the pattern was symmetrical on the vertical midline. On each trial, subjects were first presented with a 16 × 16 matrix of black and white squares and were required to decide whether the pattern was symmetric on the vertical midline. Followed by the symmetry judgement, a 4 × 4 matrix of squares with one square highlighted in red were displayed. The location of the red-square was the to-be-remembered spatial location. This alternation continued until a variable set-size of spatial locations had been presented. Then, on the recall screen the same 4 × 4 matrix of squares was presented but with no squares highlighted in red. Subjects had to recall the spatial locations in the correct order by clicking the mouse on the appropriate squares in the matrix. There was a total of 12 trials (2 blocks of 6 trials), set-sizes ranged from 2 to 7, and each set-size occurred twice (once in each block). Scores on the advanced symmetry span task were calculated using the partial-scoring method, the number of spatial locations recalled in their correct order across all trials.

Advanced rotation span. This task required subjects to remember a series of directional arrows of varying size in alternation with a mental rotation task in which they had to mentally rotate and decide if a letter was mirror reversed or not. On each trial, subjects first solved a mental rotation problem followed by the presentation of a single arrow with a specific direction (8 possible directions; the four cardinal and four ordinal directions) and specific size (small or large). Both the direction and size of the arrow were the to-be-remembered features. This alternation continued until a variable set-size of arrows had been presented. Then, on the recall screen all possible arrow directions and sizes were presented. Subjects had to recall the direction and size of the arrows in the correct order by clicking the mouse on the appropriate arrow. There was a total of 12 trials (2 blocks of 6 trials), set-sizes ranged from 2 to 7, and each set-size occurred twice (once in each block). Scores on the advanced rotation span task were calculated using the partial-scoring method, the number of arrows recalled in their correct order across all trials.

2.1.6. Attention control tasks

Antisaccade (Hallett, 1978; Hutchison, 2007). In this task, subjects had to identify a "Q" or "O" that appeared very briefly on the opposite side of the screen as a distractor stimulus. Subjects were first presented with a fixation cross at the center of the screen. After a 2000 ms or 3000 ms interval an asterisk (*) flashed at 12.3° visual angle to the left or right of the central fixation for 100 ms. After the presentation of the asterisk, either the letter "Q" or "O" was presented on the opposite side at 12.3° visual angle of the central fixation for 100 ms quickly followed by a visual mask (##). Subjects had to indicate whether the letter was a "Q" or an "O". The subject's goal was to ignore the asterisk and instead look away to the other side of the screen to catch the target "Q" or "O". Subjects had as much time as needed to respond to which letter appeared by pressing the associated key on the keyboard. After responding, accuracy feedback was displayed for 500 ms, followed by a blank inter-trial interval of 1000 ms. Subjects completed 16 slow practice trials (target duration was set to 750 ms). There were 72 real trials and scores on the task were calculated as the proportion of correct trials.

Selective visual arrays with orientation judgement – VAorient-S (Draheim et al., 2020; Luck & Vogel, 1997; Shipstead, Lindsey, Marshall, & Engle, 2014). In this task subjects had to decide whether a probe array of stimuli was the same or different from the target array. The stimuli were red (RGB: 255, 0 0) and blue (RGB: 0, 0, 255) rectangles in various orientations (horizontal, left diagonal, right diagonal, or vertical). Specifically, they had to make a judgement as to whether a single rectangle in the probe array had remained in the same orientation or was in a different orientation as the target array. They responded by pressing the 5 and 6 keys on the numpad, labeled "Yes" and "No" respectively. Prior to each trial subjects were reminded to respond "Yes" for a same judgement and "No" for a different judgement and had to press the spacebar for the trial to begin. Thus, the task was self-paced. After pressing the spacebar, there was a 1 s blank screen followed by a screen with a central fixation (+) for 1 s. After the fixation, a cue was presented, "RED" or "BLUE", to instruct the subject to attend to either red or blue rectangles, this was followed by a blank screen for 100 ms. Next, a target array of blue and red rectangles differing in orientation (horizontal, left diagonal, right diagonal, or vertical) were presented for 250 ms. After a delay (blank screen) of 900 ms, a probe array with only the color of rectangles they were cued to attend to was presented with one of the rectangles highlighted by a white dot. The rectangle with the white dot changed orientation on 50% of the trials and remained the same on the other 50% of trials, while all other rectangles were identical to the target array. The probe array remained on screen until a response was made. The subject had to make a response as to whether the rectangle remained in the same orientation, "Yes", or was in a different orientation, "No". The background color was set to "silver" (RGB: 192, 192, 192) and all words and fixation crosses were in black. The target array contained either 5 or 7 rectangles per color (10 and 14 total), and a total of 48 trials were presented for each array set size. The dependent variable was a capacity score (k), which is calculated using the single probe correction (Conway et al., 2005; Shipstead et al., 2014). This calculation is N * (hits + correction rejections – 1), where N is the set size for that array. This calculation results in two separate k scores, one for set size 5 and one for set size 7, and the final dependent variable was the average k for these two set sizes.

Sustained attention-to-cue task – SACT (Draheim et al., 2020). In this task, subjects needed to sustain their attention on a visual circle cue presented at random locations on the screen and ultimately identify a target letter presented briefly at the center of the cue. This task was designed as an accuracy version of the psychomotor vigilance task (Dinges & Powell, 1985), with the addition of a distractor similar to that used in the antisaccade task. The stimuli for the task were presented against a gray background. Each trial started with a central black fixation. On half of the trials, the fixation was presented for 2 s and for the other half the fixation was presented for 3 s. After the fixation, following a 300 ms tone, a large white circle cue was presented in a randomly determined location on either the left or right side of the screen. To orient the subject on the circle cue, the large circle began to immediately shrink in size until it reached a fixed size. Once the cue reached the fixed size, after a variable wait time (equally distributed amongst 2, 4, 8, and 12 s), a white asterisk meant to serve as a distractor appeared at the center of the screen. The asterisk blinked on and off in 100 ms intervals for a total duration of 400 ms. Then, a 3 × 3 array of letters was displayed at the center of the cue. The letters in the array consisted of B, D, P, and R. The central letter was the target letter and was presented in a dark gray font. The non-target letters were presented in black font with each letter occurring twice in the array and the target letter occurred three times. After 125 ms the central letter was masked with a # for 1000 ms. Only the central target letter was masked. After the mask, the response options were displayed in boxes horizontally across the upper half of the screen. The subject used the mouse to select whether the target was a B, D, P, or R. Feedback was given during the practice trials but not the experimental trials. Accuracy rate was the dependent variable.

2.1.7. Data processing and analysis

Data, scripts, and results outputs are open access and available at Open Science Framework: https://osf.io/ajm4d/. All preprocessing, data cleaning and scoring, and data analyses were conducted in R statistical software (R Core Team, 2020). Univariate outliers and problematic scores were removed prior to analysis, as part of data cleaning. The procedures for removing data for each measure are described below.

Missing data may be present due to data cleaning but also to other factors such as a subject not having enough time to complete a task on a given session, and the task program crashing during administration.

Baseline pupil. Preprocessing methods were employed on the raw baseline pupil data using the *pupillometry* package (Tsukahara, 2019). Only data from the left eye was preprocessed and further analyzed. First, raw pupil data were de-blinked and values 75 ms before and after blinks were set to missing. Next, raw pupil data were smoothed with a hanning filter and then linear interpolation of missing values was applied. Missing data gaps of more than 1000 ms were not interpolated. If a subject had more than 50% of missing data, then their baseline pupil data was removed from further analysis. Univariate outliers on baseline pupil size and baseline pupil variability were identified and removed separately for each of the eight baseline conditions. Outliers were identified as having values +/− 3.3 standard deviations from the mean value, within baseline condition, and outlier values were replaced with missing data.

Cognitive tasks. For all the cognitive tasks, univariate outliers were identified as having scores +/− 3.3 standard deviations or greater from the mean score on that task and outlier scores were replaced with missing data. For the visual arrays task (VAorient-S), subjects that had an overall accuracy of −3.3 standard deviations or greater from the mean accuracy had their calculated k score replaced with missing data. For the working memory capacity tasks, subjects that had accuracy on the processing portion of the task −3.3 standard deviations or greater from the mean had their calculated partial span score replaced with missing data. For the composite factors of fluid intelligence, working memory capacity, and attention control, if a subject had missing data from two or more task (out of a total of three) that make up that composite their composite score was replaced with a missing value.

Reliability estimates. A split-half reliability method was used to estimate reliability for each task. The tasks were split into even/odd trials and the scores were calculated for each half just as they were for the whole task. Correlations between even and odd scores were corrected with the spearman-brown prophecy formula. For the baseline pupil size and variability measures, reliability was estimated by correlating the measures for the first 30 s and the last 30 s and applying the spearman-brown prophecy formula.

Data analysis. To test for the within-subject effects of lighting conditions, the between-subject effects of cognitive ability, and their interactions we used hierarchical linear modelling with the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). To further investigate any significant interactions, we conducted simple slopes analysis using the *reghelper* package (Hughes, 2020). We made plots using the *ggplot2* package (Wickham, 2016) and raincloud plots (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019).

2.2. Results

2.2.1. Effect of background color on baseline pupil size

Descriptive statistics for each task and background color condition are presented in Table 2. Distributions on baseline pupil size in the gray and white background conditions are displayed in Fig. 1 (for method of generating raincloud plots see; Allen et al., 2019). For the gray background condition, the mean ($M = 5.00$) and inter-individual standard deviation ($SD = 0.82$) baseline pupil size was larger than Unsworth et al. (2019) ($M = 3.21$, $SD = 0.49$) and closer to Tsukahara et al. (2016); $M = 5.99$, $SD = 1.09$. Whereas, for the white background condition, the mean ($M = 3.64$) and inter-individual standard deviation ($SD = 0.56$) was closer to the values reported in Unsworth et al. (2019); $M = 3.21$, $SD = 0.49$. A paired-samples t-test was conducted to test for the difference in mean pupil size between the two conditions; mean pupil size in the gray background condition was 1.39 mm [95% CI: 1.34, 1.45] larger than in the white background condition, $t(286) = 48.93$, $p < .05$.

2.2.2. Background Color x Cognitive Ability interactions

Fluid intelligence, working memory capacity, and attention control composites were created. The composites were created by averaging the standardized z-scores for each task. For each composite, if two out of three of the tasks were missing, then the composite score was set to missing. Fluid intelligence correlated with working memory capacity ($r = 0.51$) and attention control ($r = 0.50$). Working memory capacity correlated with attention control ($r = 0.45$). The correlations between the baseline pupil measures and cognitive abilities are reported in Table 3 (see Fig. S4 for corresponding scatterplots and Table S1 for the correlations amongst the baseline pupil measures).

Fluid intelligence correlated with baseline pupil size in both the gray ($r = 0.29$, $p < .05$) and white ($r = 0.16$, $p < .05$) background conditions. Working memory capacity correlated with baseline pupil size in the gray ($r = 0.21$, $p < .05$) but not the white ($r = 0.09$, $p > .05$) background condition. Attention control correlated with baseline pupil size in the gray ($r = 0.25$, $p < .05$) but not the white ($r = 0.09$, $p > .05$) background condition.

For each of the cognitive abilities, a hierarchical linear model was conducted to compare the interaction between background color and cognitive ability.[4] This allowed us to test for the difference in correlation between conditions while accounting for the dependencies (within-subject nature) of pupil measurements. The results of the hierarchical linear models are presented in Tables 4–6.

There was a significant Background Color x Fluid Intelligence interaction on baseline pupil size; $\beta = 0.16$ [95% CI: 0.11, 0.22], $p < .05$. The relationship between baseline pupil size and fluid intelligence was stronger in the gray background condition than the white background condition. Simple slopes analysis indicated that the slope for the gray background condition was positive and statistically significant ($b = 0.29$, $\beta = 0.29$, $t = 6.09$, $p < .05$) whereas the slope for the white background condition was smaller but still statistically significant ($b = 0.10$, $\beta = 0.14$, $t = 2.04$, $p < .05$). There was no significant interaction on intra-individual baseline pupil variability, $\beta = -0.11$ [95% CI: −0.25, 0.04], $p > .05$, but there was a main effect of fluid intelligence on intra-individual baseline pupil variability, $\beta = 0.19$ [95% CI: 0.07, 0.30], $p < .05$.

There was a significant Background Color x Working Memory Capacity interaction on baseline pupil size; $\beta = 0.11$ [95% CI: 0.06, 0.17], $p < .05$. The relationship between baseline pupil size and working memory capacity was stronger in the gray background condition than the white background condition. Simple slopes analysis indicated that the slope for the gray background condition was positive and statistically significant ($b = 0.20$, $\beta = 0.21$, $t = 4.21$, $p < .05$) whereas the slope for the white background condition was non-significant ($b = 0.07$, $\beta = 0.10$, $t = 1.37$, $p > .05$). There was a significant Background Condition x Working Memory Capacity interaction on intra-individual baseline pupil variability; $\beta = -0.15$ [95% CI: −0.30, −0.01], $p < .05$. Simple slopes analysis revealed that the relationship between intra-individual baseline pupil variability and working memory capacity was negative and non-significant in the gray background condition ($b = -0.01$, $\beta = -0.05$, $t = -0.92$, $p > .05$) but positive and non-significant in the white background condition ($b = 0.01$, $\beta = 0.11$, $t = 1.73$, $p > .05$).

There was a significant Background Color x Attention Control interaction on baseline pupil size; $\beta = 0.15$ [95% CI: 0.09, 0.20], $p < .05$. The relationship between baseline pupil size and attention control was stronger in the gray background condition than the white background condition. Simple slopes analysis indicated that the slope for the gray background condition was positive and statistically significant ($b = 0.25$, $\beta = 0.23$, $t = 4.80$, $p < .05$) and the slope for the white background

[4] For each cognitive ability, a random intercepts model was conducted in which pupil measurements were nested within subjects, background color was a level 1 predictor, the cognitive ability was a level 2 predictor, and the interaction between background color and the ability was specified.

Table 2
Descriptive statistics for study 1 ($N = 315$).

	Mean (SD)	Min–Max	Skewness	Kurtosis	Reliability	Missing
Gf *(ACC)*						
RAPM	9.9 (3.3)	1–18	−0.24	−0.41	0.80	0.0%
LetterSets	16.8 (4.4)	5–28	−0.27	−0.48	0.86	1.0%
NumberSeries	9.6 (3.2)	2–15	−0.31	−0.77	0.85	0.3%
WMC *(ACC)*						
SymSpan	27.9 (10.4)	2–54	−0.06	−0.40	0.79	3.2%
OSpan	55.5 (14.7)	12–82	−0.61	−0.16	0.64	3.5%
RotSpan	24.4 (9.3)	1–49	0.05	−0.31	0.81	1.0%
Attention						
Antisaccade *(ACC)*	0.79 (0.15)	0.36–1.0	−0.93	0.06	0.91	2.9%
VAorient-S *(k)*	1.76 (1.21)	−1.30–5.32	0.05	−0.32	0.74	1.3%
SACT *(ACC)*	0.70 (0.19)	0.19–0.98	−0.74	−0.15	0.93	3.8%
Pupil size						
Gray (mm)	5.0 (0.82)	2.87–7.26	−0.13	−0.52	0.98	1.0%
White (mm)	3.64 (0.56)	2.40–5.91	0.70	0.90	0.78	7.9%
Pupil variability						
Gray (mm)	0.29 (0.12)	0.12–0.83	1.28	2.45	0.79	1.0%
White (mm)	0.30 (0.11)	0.06–0.64	0.50	−0.05	0.96	7.9%

Note. Gf = fluid intelligence, WMC = working memory capacity. ACC = accuracy (Gf tasks were calculated as total correct and WMC tasks using the partial scoring method. Antisaccade and SACT were calculated as proportion correct. VAorient-S was calculated as *k*). Baseline pupil size was calculated as the average pupil size over the baseline period (mean) and intra-individual baseline pupil variability was calculated as the standard deviation of pupil size over the baseline period, reported pupil values are in millimeters (mm). Reliability estimates were calculated as split-half reliability corrected with spearman-brown prophecy formula.

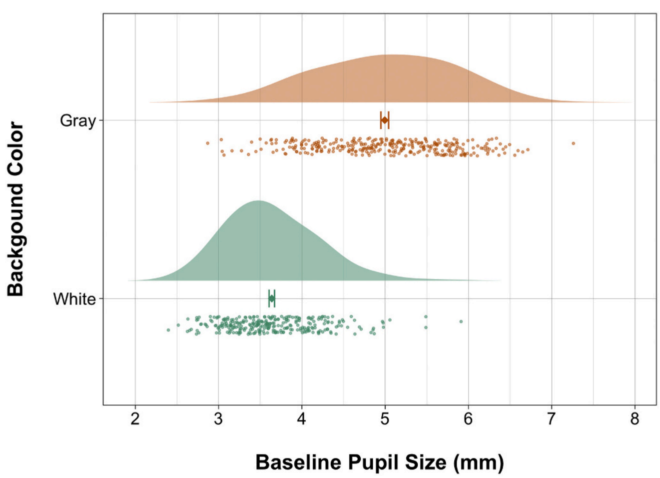

Fig. 1. Distributions of baseline pupil size values in the gray and white background conditions. The mean and 95% confidence intervals are plotted below the cloud distribution and the individual data points are at the bottom.

Table 3
Correlations between baseline pupil measures and cognitive abilities.

	Pupil size		Pupil variability	
	Gray	White	Gray	White
Fluid intelligence	**0.29**	**0.16**	0.08	**0.20**
Working memory capacity	**0.21**	0.09	−0.05	0.10
Attention control	**0.25**	0.09	−0.07	0.02

Note. Correlations were calculated using pearson-method with pairwise deletion and values in bold font are statistically significant at $p < .05$.

condition was smaller and non-significant ($b = 0.06$, $\beta = 0.08$, $t = 1.13$, $p > .05$). There was no significant interaction, $\beta = -0.10$ [95% CI: -0.24, 0.05], $p > .05$, nor a main effect of attention control on intra-individual baseline pupil variability, $\beta = 0.02$ [95% CI: -0.10, 0.13], $p > .05$.

2.2.3. *Fluid intelligence, working memory capacity, or attention control?*

Next, we conducted a regression analysis to test the unique relationships of fluid intelligence, working memory capacity, and attention control on baseline pupil size in the gray background condition only. In Tsukahara et al. (2016) we found that fluid intelligence, not working memory capacity, uniquely predicted baseline pupil size. Again, in this study, we found that only fluid intelligence ($\beta = 0.19$ [95% CI: 0.05, 0.32], $p < .05$), and neither working memory capacity ($\beta = 0.07$ [95% CI: -0.06, 0.20], $p > .05$) nor attention control ($\beta = 0.11$ [95% CI: -0.01, 0.24], $p > .05$), uniquely predicted baseline pupil size.

2.2.4. *Age*

Winn et al. (1994) found that baseline pupil size decreased with age, however, this correlation was reduced with brighter luminance. Our hypothesis was that brighter luminance will reduce the range of baseline pupil size values and thereby reduce the correlation with cognitive abilities. This hypothesis was informed by the findings in Winn et al. (1994) and therefore we expected to see the same interaction between background color and age on baseline pupil size as we did with cognitive ability.

As expected, we replicated the findings in Winn et al. (1994). There was a significant Background Color x Age interaction on baseline pupil size; $\beta = -0.19$ [95% CI: -0.24, -0.13], $p < .05$. Simple slopes analysis indicated that the slope for the gray background condition was positive and statistically significant ($b = -0.06$, $\beta = -0.28$, $t = -5.73$, $p < .05$) and the slope for the white background condition was smaller and non-significant ($b = -0.01$, $\beta = -0.09$, $t = -1.24$, $p > .05$). There was no significant interaction on intra-individual baseline pupil variability, $\beta = 0.05$ [95% CI: -0.10, 0.19], $p > .05$, but there was a main effect of age on intra-individual baseline pupil variability, $\beta = -0.19$ [95% CI: -0.30, -0.07], $p < .05$.

Even though our age range was only 18–35, it is important to control for age as a potential confounding factor in the relationship between pupil size and cognitive ability. In our sample, age did correlate with baseline pupil size ($r = -0.28$), fluid intelligence ($r = -0.34$), working memory capacity ($r = -0.20$) and attention control ($r = -0.16$). For each cognitive ability, a linear regression was conducted in which we included age and cognitive ability as a predictor of baseline pupil size in the gray background condition only. After controlling for age, fluid intelligence still predicted baseline pupil size, $\beta = 0.22$ [95% CI: 0.10, 0.33], $p < .05$. The same was true of working memory capacity, $\beta = 0.16$ [95% CI: 0.05, 0.27], $p < .05$, and attention control, $\beta = 0.21$ [95% CI: 0.10, 0.32], $p < .05$.

Table 4
Hierarchical linear model: Background Color x Fluid Intelligence (Gf).

Predictors	Baseline pupil size			Baseline pupil variability		
	B	*95% CI*	*p*	*B*	*95% CI*	*p*
Intercept	−0.74	−0.82 to −0.66	**<0.001**	0.02	−0.09–0.13	**<0.001**
Background Color	1.42	1.36–1.47	**<0.001**	−0.03	−0.17–0.11	0.683
Gf	0.08	0.00–0.16	**0.042**	0.19	0.07–0.30	**0.001**
Background Color x Gf	0.16	0.11–0.22	**<0.001**	−0.11	−0.25–0.04	0.146
Random effects						
σ^2			0.10			0.01
τ_{00}			0.38 $_{\text{Subject}}$			0.00 $_{\text{Subject}}$
ICC			0.78			0.21
N			315 $_{\text{Subject}}$			315 $_{\text{Subject}}$
Observations			602			602
Marginal R^2 / Conditional R^2			0.517 / 0.895			0.020 / 0.223

Table 5
Hierarchical linear model: Background Color x Working Memory Capacity (WMC).

Predictors	Baseline pupil size			Baseline pupil variability		
	B	*95% CI*	*p*	*B*	*95% CI*	*p*
Intercept	−0.74	−0.82 to −0.65	**<0.001**	0.02	−0.10–0.14	**<0.001**
Background Color	1.42	1.36–1.47	**<0.001**	−0.03	−0.17–0.11	0.698
WMC	0.06	−0.02–0.14	0.170	0.10	−0.01–0.22	0.084
Background Color x WMC	0.11	0.06–0.17	**<0.001**	−0.15	−0.30 to −0.01	**0.034**
Random effects						
σ^2			0.11			0.01
τ_{00}			0.38 $_{\text{Subject}}$			0.00 $_{\text{Subject}}$
ICC			0.77			0.22
N			311 $_{\text{Subject}}$			311 $_{\text{Subject}}$
Observations			595			595
Marginal R^2 / Conditional R^2			0.502 / 0.887			0.007 / 0.228

Table 6
Hierarchical linear model: Background Color x Attention Control (AC).

Predictors	Baseline pupil size			Baseline pupil variability		
	B	*95% CI*	*p*	*B*	*95% CI*	*p*
Intercept	−0.73	−0.81 to −0.65	**<0.001**	0.02	−0.10–0.13	**<0.001**
Background Color	1.41	1.36–1.47	**<0.001**	−0.03	−0.17–0.12	0.721
AC	0.05	−0.03–0.13	0.258	0.02	−0.10–0.13	0.787
Background Color x AC	0.15	0.09–0.20	**<0.001**	−0.10	−0.24–0.05	0.184
Random effects						
σ^2			0.11			0.01
τ_{00}			0.38_{Subject}			0.00_{Subject}
ICC			0.78			0.22
N			313_{Subject}			313_{Subject}
Observations			599			599
Marginal R^2 / Conditional R^2			0.504 / 0.891			0.004 / 0.226

2.3. Discussion

Study 1 was successful in the attempt to reduce the correlation between baseline pupil size and cognitive abilities by increasing the luminance of the baseline conditions such that baseline pupil size values become restricted in range. The baseline pupil values in the white background color condition were similar, though still slightly larger, to that reported in Unsworth et al. (2019); see Table 2 and Fig. 1. However, the baseline pupil size values in the gray background condition were more similar to that reported in Tsukahara et al. (2016). Based on the distributions of pupil size in the two conditions, we expected a significant correlation in the gray background condition but not the white background condition. With regard to working memory capacity, this is exactly what we found; working memory capacity correlated with baseline pupil size in the gray ($r = 0.21$, $p \leq .05$) but not the white ($r = 0.09$, $p \geq .05$) background condition. In terms of numerical values, the correlation reported in Unsworth et al. (2019) was smaller, $r = 0.01$, compared to the current study of $r = 0.09$. However, we can be confident in the result from the current study due to the larger and more diverse sample. Similarly, the baseline pupil size – fluid intelligence and attention control relationships were reduced in the white background condition compared to the gray background condition. Overall, these results suggest that too bright of lighting conditions will attenuate the correlation of baseline pupil size with cognitive abilities due to restrictions of range on observed pupil values.

We also replicated the finding from Tsukahara et al. (2016) that fluid intelligence, and not working memory capacity, uniquely predicted baseline pupil size. Overall, cognitive abilities only explained a total of 9.2% of variance in baseline pupil size – this is not a large effect. Fluid intelligence and common variance (common to the cognitive abilities) explained 80% of that 9.2% of variance in baseline pupil size. Whereas, working memory capacity and common variance explains 49% of the 9.2% of variance. We would argue that the much larger contribution by fluid intelligence and common variance is a reason the baseline pupil size – fluid intelligence relationship is more robust than that with working memory capacity.

Curiously, although the gray background condition was an attempt to measure baseline pupil size under similar lighting conditions as Unsworth et al. (2019), we only saw a restriction of range in pupil values in the white background condition. This suggested to us that there is some source of illumination that we, nor other researchers, have taken into account when measuring pupil size. Therefore, we conducted another study focused more on manipulating various sources of luminance.

3. Study 2

One problem in resolving the issues with different baseline pupil size values across studies is how lighting conditions are measured. We believe there is at least one other lighting factor that can contribute to luminance of the testing environment – one that most researchers do not report; the brightness and contrast settings on the monitor (not just the background color). Besides brightness and contrast settings on the monitor, the overall brightness will vary widely from one type of monitor to the next. Given this, perhaps a more meaningful measure of luminance would take into account at least three factors; room lighting conditions, background/stimuli color, and brightness of the monitor. Although the last two factors essentially provide the same source of luminance (monitor brightness), they can be manipulated independently of one another.

Based on our findings from Study 1, we surmise that the monitor brightness/contrast used in the Unsworth lab might be considerably higher than used in our lab. This would explain why in Study 1 our baseline pupil values in the gray background condition (equivalent background as Unsworth et al., 2019) were larger, more variable, and correlated with fluid intelligence, working memory capacity, and attention control, whereas Unsworth et al. (2019) reported small and non-significant correlations. Their monitor brightness was likely more equivalent to our white background condition.

Given that the monitor brightness is ubiquitously ignored in most pupillometry research, we conducted a second study in which we investigated the impact of all three factors (room lights, background color, monitor brightness) on baseline pupil values and their relationship to cognitive abilities. We also took more precise measurements of luminance from the overall room lights and the monitor. Based on our results and these measurements of luminance we provide strong recommendations to pupillometry researchers as to the best way to report lux values.

3.1. Method

3.1.1. Subjects

College and non-college adults of the Atlanta community participated and were required to be native English speakers, 18–35 years of age, and had not participated in a study with our lab before. Screening on vision was not performed. The study consisted of two 2-h sessions. Subjects were compensated with an average of $35 on each session or receive 2 h course credit for any session instead of monetary compensation. The study was approved by the Georgia Institute of Technology's Institutional Review Board under Protocol H19016. Baseline pupil measures were obtained from a total of 201 subjects. Demographic information for this sample of subjects is presented in Table 7.

Table 7
Subject demographic for study 2 ($N = 201$).

Demographic	Category	Value
Age (Years)	Mean	22.7
	SD	4.4
Gender	Male	53%
	Female	47%
Education	Some high school	< 1%
	High school/GED	2%
	Some college	66%
	Associates degree	4%
	Bachelor's degree	20%
	Some graduate school	3%
	Master's degree	5%
	PhD/MD/JD/DDS	< 1%
Ethnicity[a]	White	38%
	Black or African American	21%
	Asian or Pacific Islander	28%
	Hispanic or Latino	12%
	Native American	<1%
	Other	<1%

[a] Other includes mixed race and other.

3.1.2. Tasks and procedures

Testing was conducted in a group running room with a total of three subject stations and one research assistant that monitored subjects and administered tasks. The tasks and baseline measures were conducted on a Windows computer with an LED-backlit LCD monitor and subjects wore headphones during all tasks and baseline measures. The tasks were programmed in *E*-Prime 3.0 software (Psychology Software Tools, 2016).

3.1.3. Baseline pupil measures

A SensoMotoric Instruments Red250m eye-tracker was used to record binocularly at 250 Hz. Subjects were seated approximately 65–70 cm from the monitor and did not use any head immobilization device. Baseline pupil size was measured at the start of both sessions. We manipulated the room lights, background color, and brightness/contrast settings on the monitor. Room lights were either on or off. The background color was either white or black. The brightness/contrast setting was either bright or dim. This resulted in a 2 × 2 × 2 within-subjects design. Half of the subjects had the room lights on for the first session and off for the second session, and the other half had the reverse order. Within each session there were four counterbalanced conditions (see Table S2).

Because brightness/contrast setting had to be manipulated manually on the monitor, and this required removing the eye-tracker from the monitor, we kept the brightness setting at 80% and set the contrast to 20% (dim condition) and 100% (bright condition). The luminance values (*lux*) are reported in Table 8. A Sper Scientific Direct Light Meter Lux – 840,006 was used to measure luminance.

Baseline pupil size was measured in four conditions per session for a total of eight conditions. Subjects were instructed that they did not have to do anything in particular during the baseline and to keep their gaze towards the monitor. Each baseline condition lasted for 1 min and there was a break between each condition to allow the research assistant to manually adjust the contrast setting on the monitor. After the monitor setting was adjusted, a screen appeared displaying the position of the subject's eyes (white ovals) in a black box. This indicated whether the eye-tracker was able to detect the eyes. The subject was instructed to position themselves so that their eyes (white ovals) were centered within the black box. Once the research assistant verified their position the subject was instructed to press the spacebar when they were ready to continue onto the next baseline condition. There were an extra 5 s added before baseline pupil recording started to allow for the pupil to adjust to the change in lighting condition. An eye-tracking calibration procedure was conducted immediately before the first condition in each session. A validation procedure was conducted immediately before the third

Table 8
Lux values in the different lighting conditions measured from the room lights, screen lighting, and participant view.

Room lighting				
	Room Lights: Off		Room Lights: On	
	1		253	
Screen lighting				
	Background Color			
	Black		White	
Setting: Dim	1		15	
Setting: Bright	8		208	
Participant view				
	Room Lights: Off		Room Lights: On	
	Background Color			
	Black	White	Black	White
Setting: Dim	1	15	43	47
Setting: Bright	1	27	44	83

Note. The values reported here are the average values obtained from three different subject stations. Room lighting was measured by placing the light meter on top of the subject station table faced up and with the monitor turned off. Screen lighting was measured by placing the light meter directly on the face of the monitor screen in the center (to capture any luminance due to the fixation). Participant view was measured by placing the light meter in a position similar to a typical subject's eye distance (60–70 cm) and angle relative to the monitor.

condition to ensure quality data. The baseline procedure for each session lasted approximately 5–10 min.[5] Two measures were obtained from preprocessed pupil data. Baseline pupil size was calculated as the average pupil size over the baseline period. Intra-individual baseline pupil variability was calculated as the standard deviation of pupil size over the baseline period.

We also obtained subjective reports of arousal after each baseline period using the same arousal scale as in (Åkerstedt & Gillberg, 1990) and (Stawarczyk & D'Argembeau, 2016). Immediately after each baseline period the arousal scale appeared on the screen asking subjects to "Rate your level of arousal using the scale below". A scale was presented as a sequence of numbers, 1–9, from left to right. Some of the numbers also had a descriptive label underneath; 1 = "Extremely Alert", 3 = "Alert", 5 = "Neither Alert nor Sleepy", 7 = "Sleepy – but no difficulty remaining awake", 9 = "Extremely sleepy – fighting sleep". In all data analyses, arousal responses were reverse scored so that higher values represent higher arousal. Overall, most of the effects of arousal were non-significant and are reported in Supplemental Materials.

3.1.4. Cognitive tasks

We measured fluid intelligence with the *Raven's Advanced Progressive Matrices*, *letter sets*, and *number series*. Working memory capacity was measured with the *advanced versions* of the *operation span*, *symmetry span*, and *rotation span* tasks. Attention control was measured with the *antisaccade*, *selective visual arrays*, and *sustained attention to cue task*. The procedures of the tasks were identical to that in Study 1.

3.1.5. Data processing and analysis

Data, scripts, and results outputs are open access and available at Open Science Framework: https://osf.io/ajm4d/. All preprocessing, data cleaning and scoring, and data analyses were conducted in R statistical software (R Core Team, 2020). Univariate outliers and problematic scores were removed prior to analysis, as part of data cleaning. The procedures for removing data for each measure are described below. Missing data may be present due to data cleaning but also to other factors such as a subject not having enough time to complete a task on a given session, and the task program crashing during administration.

Baseline pupil Preprocessing methods were employed on the raw baseline pupil data using the *pupillometry* package (Tsukahara, 2019). Only data from the left eye was preprocessed and further analyzed. First, raw pupil data were de-blinked and values 75 ms before and after blinks were set to missing. Next, raw pupil data were smoothed with a hanning filter and then linear interpolation of missing values was applied. Missing data gaps of more than 1000 ms were not interpolated. If a subject had more than 50% of missing data, then their baseline pupil data was removed from further analysis. Univariate outliers on baseline pupil size and baseline pupil variability were identified and removed separately for each of the eight baseline conditions. Outliers were identified as having values +/− 3.3 standard deviations from the mean value, within baseline condition, and outlier values were replaced with missing data. Single outliers were only present in the two conditions with background color white and monitor setting bright (the results did not change whether we kept or removed those outliers).

Cognitive tasks. For all the cognitive tasks, univariate outliers were identified as having scores +/− 3.3 standard deviations or greater from the mean score on that task and outlier scores were replaced with missing data. For the visual arrays task (VAorient-S), subjects that had an overall accuracy of −3.3 standard deviations or greater from the mean accuracy had their calculated *k* score replaced with missing data. For the working memory capacity tasks, subjects that had accuracy on the processing portion of the task −3.3 standard deviations or greater from the mean had their calculated partial score replaced with missing data. For the composite factors of fluid intelligence, working memory capacity, and attention control, if a subject had missing data from two or more task (out of a total of three) that make up that composite their composite score was replaced with a missing value.

Reliability estimates. A split-half reliability method was used to estimate reliability for each task. The tasks were split into even/odd trials and the scores were calculated for each half just as they were for the whole task. Correlations between even and odd scores were corrected with the spearman-brown prophecy formula. For the baseline pupil size and variability measures, reliability was estimated by correlating the measures for the first 30 s and the last 30 s and applying the spearman-brown prophecy formula. Because arousal for each baseline condition was not an aggregate score, reliability for arousal had to be calculated as a single estimate across the eight conditions using the same even/odd split-half reliability method as the tasks (the eight conditions were split into even/odd depending on the counterbalance order for the subject).

Data analysis. To test for the within-subject effects of lighting conditions, the between-subject effects of cognitive ability, and their interactions we used hierarchical linear modelling with the *lme4* package (Bates et al., 2015). To further investigate any significant interactions, we conducted simple slopes analysis using the *reghelper* package (Hughes, 2020). We used the *lavaan* package (Rosseel, 2012) to conduct structural equation models. We plotted the simple slopes analysis using the *sjPlot* package (Lüdecke, 2020), and any other plots using the *ggplot2* package (Wickham, 2016) and raincloud plots (Allen et al., 2019).

3.2. Results

Descriptive statistics for each cognitive task are presented in Table 9, the mean pupil size and intra-individual variability in each baseline condition are presented in Table 10. See Appendix A for the reliability estimates of each baseline pupil measure. There are clearly very different distributions of pupil size values across the eight different

[5] When the testing room was full, up to 3 subjects at a time, a subject may potentially have to wait in between each baseline condition until the experimenter was finished adjusting the monitor settings for the other subjects. Therefore, the amount of time the entire baseline procedure lasted would vary.

Table 9
Descriptive statistics for cognitive tasks in study 2 ($N = 201$).

	Mean (SD)	Min–Max	Skewness	Kurtosis	Reliability	Missing
Fluid Intelligence (ACC)						
RAPM	10.4 (3.4)	1–18	−0.60	0.15	0.78	0.5%
LetterSets	15.5 (4.1)	4–25	−0.32	−0.26	0.89	1.0%
NumberSeries	9.9 (2.9)	1–15	−0.11	−0.48	0.80	0.5%
Working Memory Capacity (ACC)						
SymSpan	29.1 (9.0)	5–51	−0.11	−0.35	0.78	4.5%
OSpan	55.0 (15.2)	5–83	−0.71	0.41	0.83	2.5%
RotSpan	23.3 (9.2)	4–47	−0.03	−0.54	0.81	2.5%
Attention Control						
Antisaccade *(ACC)*	0.87 (0.12)	0.5–1.0	−1.4	1.3	0.89	2.5%
VAorient-S *(k)*	2.09 (1.23)	−0.62–4.45	−0.19	−0.70	0.79	0.0%
SACT *(ACC)*	0.89 (0.10)	0.53–1.00	−1.36	1.43	0.87	4.0%

Note. Gf = fluid intelligence, WMC = working memory capacity. ACC = accuracy (Gf tasks were calculated as total correct and WMC tasks using the partial scoring method. Antisaccade and SACT were calculated as proportion correct. VAorient-S was calculated as *k*). Reliability estimates were calculated as split-half reliability corrected with spearman-brown prophecy formula.

Table 10
Baseline pupil size and variability in the eight baseline conditions (N = 201).

	Room Lights: Off		Room Lights: On	
	Background color			
	Black	White	Black	White
Baseline pupil size (mm)				
Setting: Dim	6.11 (0.85)	4.40 (0.82)	5.25 (0.84)	4.51 (0.80)
Setting: Bright	5.68 (0.89)	**3.07 (0.34)**	4.95 (0.80)	**3.18 (0.35)**
Baseline pupil variability (mm)				
Setting: Dim	0.27 (0.12)	0.31 (0.11)	0.24 (0.09)	0.25 (0.10)
Setting: Bright	0.42 (0.14)	**0.17 (0.07)**	0.29 (0.11)	**0.17 (0.07)**

Note. Baseline pupil size was measured as average pupil size (mm) over the baseline period. Baseline pupil variability was measured as the standard deviation of pupil size (mm) over the baseline period. Reported values are group means with standard deviation in parentheses. The two brightest conditions are in bold font.

lighting conditions (see Fig. 2).

3.2.1. Effect of lighting condition on baseline pupil size

A hierarchical linear model was conducted to test for the effect of each factor (room lights, background color, and monitor setting) and their interactions.[6] The model results are presented in Table 11. There was a main effect of room lights, background color, and monitor setting such that mean baseline pupil size was smaller in the brighter conditions. All the two-way interactions were significant. For the Room Lights x Background Color interaction, the effect of background color (black vs. white) was larger when the room lights were off compared to when the room lights were on. This appears to be due to pupil size being larger on a black background when room lights are off compared to room lights on, however there is little difference due to room lights on a white background. For the Room Lights x Monitor Setting interaction, the effect of monitor setting (dim vs. bright) was larger when the room lights were off compared to room lights on. For the Background Color x Monitor Setting interaction, the effect of monitor setting (dim vs. bright) was smaller when the background color was black compared to a white background color. The three-way interaction was non-significant.

More importantly, the two conditions in which the background color was white and monitor setting was bright displayed particularly small mean and reduced inter-individual pupil size values (see Fig. 2). Based on the "Screen Lighting" measurement of lux values, these were also the two brightest conditions, see Table 8. In fact, these values were similar to that reported in Unsworth et al. (2019); $M = 3.21$, $SD = 0.49$. Therefore, if reduced variability in baseline pupil size due to bright lighting conditions leads to smaller correlations with cognitive abilities then we should see a reduced correlation in these two conditions compared to the other six conditions. To test this hypothesis, we examined the interaction between the lighting conditions and cognitive ability.

3.2.2. Lighting Condition x Cognitive Ability interactions

Fluid intelligence, working memory capacity, and attention control composites were created. The composites were created by averaging the standardized *z*-scores for each task. For each composite, if two out of three of the tasks were missing, then the composite score was set to missing. Fluid intelligence correlated with working memory capacity ($r = 0.54$) and attention control ($r = 0.53$). Working memory capacity correlated with attention control ($r = 0.43$).

Fluid intelligence correlated with baseline pupil size in all of the eight baseline conditions, working memory capacity correlated with baseline pupil size in one of the eight baseline conditions, and attention control correlated with seven out of eight of the baseline conditions (Table 12). It was expected that working memory capacity would at least correlate with baseline pupil size in the darker baseline conditions, however the only significant correlation was in one of the brightest conditions. This could be due to low power of a smaller sample size compared to our previous studies. Regardless, it still suggests that the baseline pupil size – working memory capacity relationship is not as robust as the relationship with fluid intelligence.

Overall, there were small and non-significant correlations between intra-individual baseline pupil variability and cognitive ability (Table S3). Although the baseline pupil size measures were all highly correlated with one another, the intra-individual baseline pupil variability measures from one condition to the next were weakly correlated (Table S4).

To test if the baseline pupil size – cognitive ability relationship changes depending on lighting conditions, we conducted hierarchical linear models separately for each cognitive ability. These models were the same as that presented in Table 11, but with the main effect and interaction terms for cognitive ability added.

Fluid intelligence. There was a Fluid Intelligence x Background Color x Monitor Setting interaction; $\beta = -0.11$ [95% CI: -0.21, -0.01], $p < .05$ (Table S5). To further investigate this interaction, we plotted the results of this model in Fig. 3 and conducted simple slopes analysis. From Fig. 3, it appears that the relationship between baseline pupil size and fluid intelligence is smaller and non-significant in the two brightest conditions compared to the other six conditions. Simple slopes analysis confirmed this; the slope for the On_White_Bright ($b = 0.07$, $\beta = 0.17$, $t = 1.18$, $p > .05$) and Off_White_Bright ($b = 0.09$, $\beta = 0.23$, $t = 1.50$, $p > .05$) conditions were non-significant and the slope for the other

[6] A random intercepts model was conducted in which mean baseline pupil measures were nested within subjects. Room lights, background color, monitor setting and their interactions were level 1 predictors.

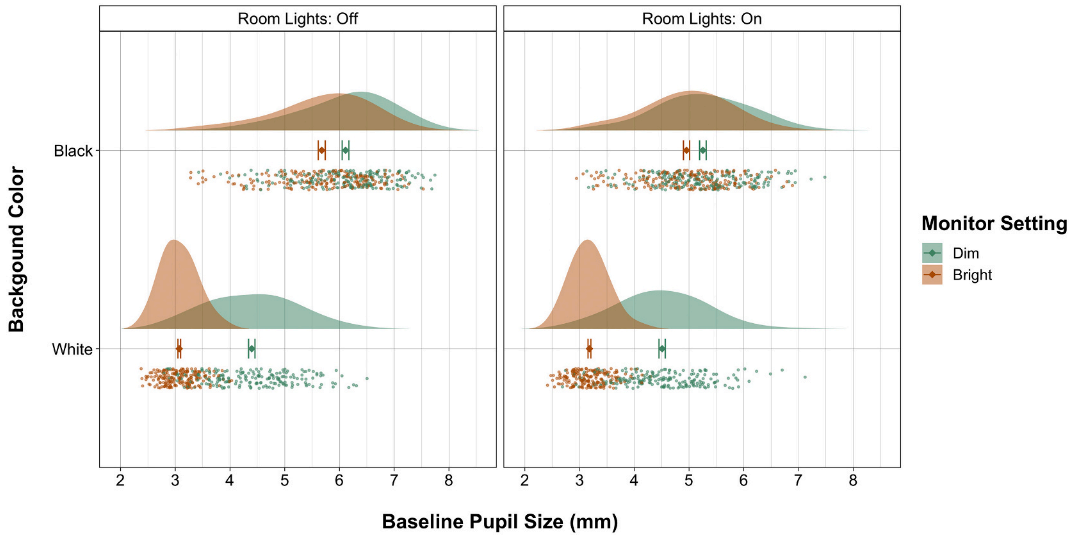

Fig. 2. Distributions of baseline pupil size values in the eight baseline conditions. Means and 95% confidence intervals are plotted below the cloud distribution, and individual data points are displayed at the bottom.

Table 11
Hierarchical linear model: Room Lights x Background Color x Monitor Setting.

Predictors	Baseline pupil size		
	B	*95% CI*	*p*
Intercept	1.17	1.08–1.25	**<0.001**
Room lights	−0.69	−0.77 to −0.62	**<0.001**
Background color	−1.38	−1.45 to −1.30	**<0.001**
Monitor setting	−0.34	−0.41 to −0.27	**<0.001**
Room Lights × Background Color	0.78	0.68–0.88	**<0.001**
Room Lights × Monitor Setting	0.12	0.02–0.21	**0.023**
Background Color × Monitor Setting	−0.68	−0.78 to −0.58	**<0.001**
Room Lights × Background Color × Monitor Setting	−0.13	−0.27–0.01	0.066
Random effects			
σ^2		0.20	
$\tau_{00\ Subject}$		0.36	
ICC		0.65	
$N_{Subject}$		200	
Observations		1546	
Marginal R^2 / conditional R^2		0.648 / 0.876	

Table 12
Correlation table between cognitive abilities and baseline pupil size (mm).

Baseline pupil size	Fluid intelligence	Working memory capacity	Attention control
Off_Black_Dim	**0.16**	0.05	**0.20**
Off_Black_Bright	**0.17**	0.07	**0.23**
On_Black_Dim	**0.20**	0.10	**0.18**
On_Black_Bright	**0.24**	0.13	**0.21**
On_White_Dim	**0.21**	0.12	**0.17**
Off_White_Dim	**0.26**	0.10	**0.19**
On_White_Bright	**0.17**	0.13	**0.18**
Off_White_Bright	**0.25**	**0.18**	**0.20**

Note. Correlations were calculated using pearson-method with pairwise deletion and values in bold font are statistically significant at $p < .05$.

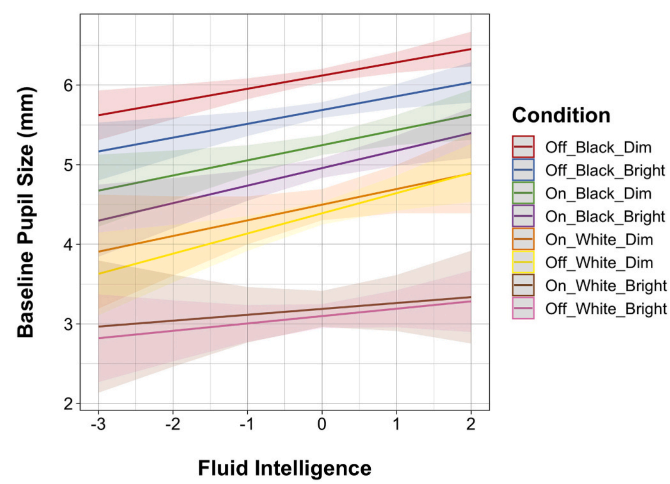

Fig. 3. Fluid intelligence x Lighting Condition interaction. The correlation of baseline pupil size and fluid intelligence was smaller and non-significant in the two brightest conditions (the bottom two slopes).

conditions were all significant. Evaluating the standardized simple slopes yields a similar pattern as seen in the correlations from Table 12; however, the Fluid Intelligence x Background Color x Monitor Setting interaction and simple slopes analysis (Fig. 3) suggest less confidence as to whether there is a relationship between baseline pupil size and fluid intelligence in the two brightest conditions only.

Working memory capacity. There was no overall effect of working memory capacity on baseline pupil size, $\beta = 0.03$ [95% CI: −0.05–0.12], $p > .05$, and there was no significant interactions between working memory capacity and the lighting conditions (Table S6). The lack of interactions is due to the weak correlations with working memory capacity (Table 12) in each of the eight conditions. Therefore, a larger sample size would likely be necessary to detect a change from a weak correlation to a weaker correlation. The results of the model are plotted in Fig. 4.

Attention control There was an Attention Control x Background Color x Monitor Setting interaction; $\beta = -0.12$ [95% CI: −0.22, −0.02], $p < .05$ (Table S7). To further investigate this interaction, we plotted the results of this model in Fig. 5 and conducted simple slopes analysis. From Fig. 5 it appears that the relationship between baseline pupil size

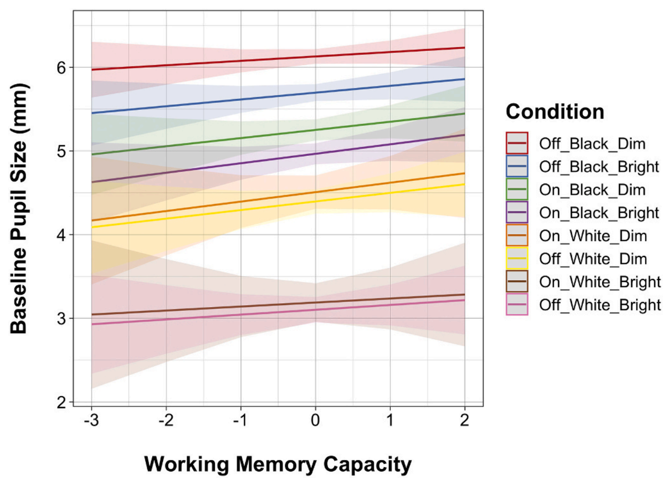

Fig. 4. Working Memory Capacity x Lighting Condition interaction. There was no significant interaction or overall effect of working memory capacity.

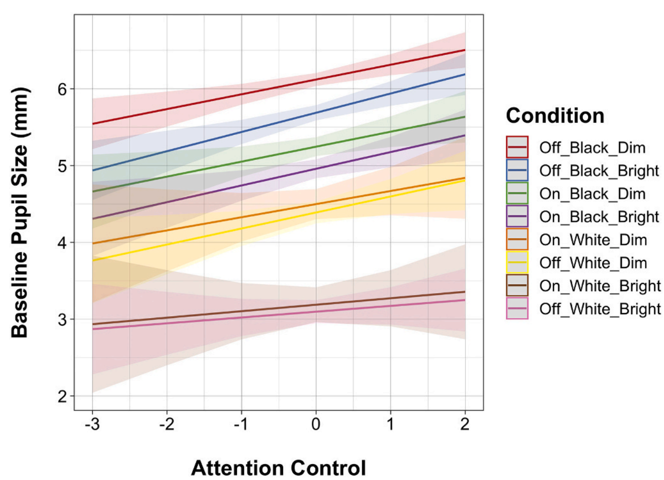

Fig. 5. Attention Control x Lighting Condition interaction. The correlation of baseline pupil size and attention control was smaller and non-significant in the two brightest conditions (the bottom two slopes).

and attention control is smaller and non-significant in the two brightest conditions compared to the other six conditions. Simple slopes analysis confirmed this; the slope for the On_White_Bright ($b = 0.08$, $\beta = 0.19$, $t = 1.26$, $p > .05$) and Off_White_Bright ($b = 0.08$, $\beta = 0.18$, $t = 1.14$, $p > .05$) conditions were non-significant and the slope for the other conditions were all significant. Evaluating the standardized simple slopes yields a similar pattern as seen in the correlations from Table 12; however, the Attention Control x Background Color x Monitor Setting interaction and simple slopes analysis (Fig. 5) suggest less confidence as to whether there is a relationship between baseline pupil size and attention control in the two brightest conditions only.

3.2.3. *Fluid intelligence, working memory capacity, or attention control?*

Next we conducted structural equation models to test the unique and common relationships of fluid intelligence, working memory capacity, and attention control on baseline pupil size (see Appendix B for the full correlation matrix). Based on the interaction results above, we excluded the On_White_Bright and Off_White_Bright from the analyses. We loaded baseline pupil size from the other six conditions onto a common latent factor. The latent Pupil Size factor, therefore, represents reliable and common variance across the baseline conditions. First, we conducted a model with Fluid Intelligence, Working Memory Capacity, and Attention Control as correlated predictors of Pupil Size. In this model, none of the cognitive abilities predicted statistically significant unique variance in Pupil Size; $\chi^2(81) = 109.08$, $p < .05$, CFI = 0.99, RMSEA [95% CI] = 0.04 [0.02, 0.06]. In terms of their relative contribution, Fluid Intelligence (~34%) and Attention Control (~45%) contributed more unique variance in Pupil Size (a total of ~9% explained variance) than Working Memory Capacity (~7%). Nevertheless, given that none of the unique paths were statistically significant this suggests that common variance across the cognitive tasks should predict Pupil Size.

To test for this, we then conducted a bi-factor model with Common variance across all the tasks predicting Pupil Size. In this model (Fig. 6), only the Common latent factor predicted variance in Pupil Size, $\beta = 0.28$ [95% CI: 0.09, 0.46], $p < .05$.

3.2.4. *Age*

We tested whether there was an interaction between age and lighting condition. Our hypothesis was that if too bright of lighting conditions restricts the range of baseline pupil size values this should reduce the correlation not only with cognitive ability but also with age. We conducted the same model as the Lighting Condition x Cognitive Ability models except with age instead of cognitive ability. Indeed, there was a significant Age x Background Color x Monitor Setting interaction on baseline pupil size; $\beta = 0.10$ [95% CI: 0.00, 0.20], $p < .05$. None of the other Age x Lighting Condition interactions were significant. Simple slopes analysis indicated that the slope for the On_White_Bright ($b = -0.03$, $\beta = -0.32$, $t = -2.17$, $p < .05$) was significant but the slope for the Off_White_Bright ($b = -0.02$, $\beta = -0.27$, $t = 1.79$, $p > .05$) condition was non-significant and the slope for the other conditions were all significant.

Next, we tested whether age can account for the baseline pupil size – cognitive ability relationship in several structural equation models. Even when Age was included as a predictor in the bi-factor model presented in Fig. 6, the Common latent factor of cognitive ability still predicted Pupil Size, $\beta = 0.21$ [95% CI: 0.05, 0.37], $p < .05$. Age also predicted Pupil Size, $\beta = -0.39$ [95% CI: -0.62, -0.17], $p < .05$. Notably, even though Age and the Common latent factor were correlated ($r = -0.23$) the path value from the Common factor to Pupil Size with ($\beta = 0.21$) and without ($\beta = 0.28$) Age were similar in magnitude.

We then performed a series of structural equation models to test how robust each of the baseline pupil size – cognitive ability relationships were when controlling for age. To do so, we first specified a structural equation model with only the cognitive ability latent factor predicting the Pupil Size latent factor, then in another model added Age as a correlated predictor of Pupil Size. We performed this series of two models (cognitive ability-only and cognitive ability-with age) separately for each cognitive ability. In the cognitive ability-only models, Fluid Intelligence ($\beta = 0.27$ [95% CI: 0.12, 0.41], $p < .05$), Working Memory Capacity ($\beta = 0.17$ [95% CI: 0.12, 0.33], $p < .05$), and Attention Control ($\beta = 0.26$ [95% CI: 0.10, 0.43], $p < .05$) significantly predicted Pupil Size. In the cognitive ability-with age models, only Fluid Intelligence ($\beta = 0.17$ [95% CI: 0.01, 0.32], $p < .05$), and Attention Control ($\beta = 0.20$ [95% CI: 0.03, 0.36], $p < .05$) significantly predicted Pupil Size after controlling for Age. Working Memory Capacity did not predict Pupil Size after controlling for Age ($\beta = 0.04$ [95% CI: -0.12, 0.21], $p > .05$). Therefore, based on these structural equation models, the baseline pupil size – working memory capacity relationship is not as robust to the potential confound of age.

3.3. *Discussion*

To summarize the results from Study 2, we showed the mean and variance of baseline pupil size values can be reduced in bright lighting conditions (Fig. 2). Specifically, it was the combination of a white background and bright monitor settings that had the largest effect. This creates a floor effect by reducing baseline pupil size down to the minimum of physiological limits. In fact, the mean and inter-individual

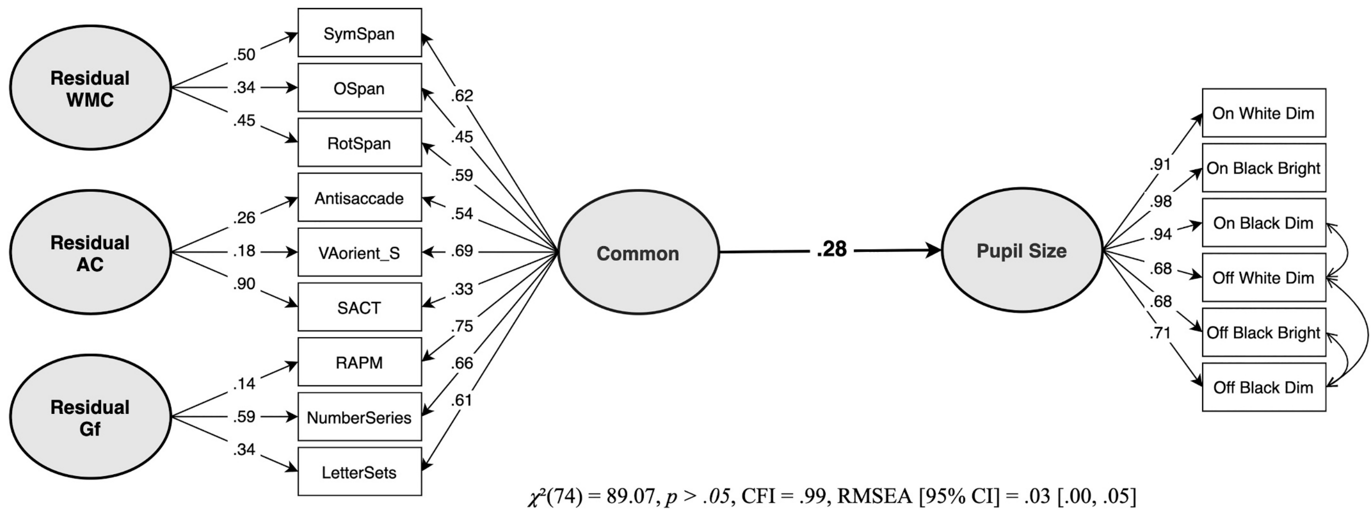

Fig. 6. Bi-factor model with a Common latent factor predicting Pupil Size. Paths from the latent Residual factors to Pupil Size were also included but were non-significant; Residual Working Memory Capacity (WMC) $\beta = -0.06$, Residual Attention Control (AC) $\beta = 0.04$, and Residual Fluid Intelligence $\beta = 0.08$.

variability of baseline pupil values in the brightest lighting conditions were very similar to that reported in Unsworth et al. (2019). More importantly, we showed fluid intelligence and attention control predicted mean baseline pupil size in every lighting condition except for the two brightest conditions in which the background monitor was white and monitor setting was set to bright. At the bivariate level, working memory capacity did not correlate with baseline pupil size in any of the lighting conditions.

In the structural equation models, we showed that common variance across the cognitive ability tasks, not unique variance from the three latent constructs, predicted baseline pupil size. This was true even after controlling for age. Additionally, at the latent construct level, only working memory capacity no longer predicted baseline pupil size after controlling for age.

Overall, these results suggest that the baseline pupil size – fluid intelligence relationship is more robust than the baseline pupil size – working memory capacity relationship. At the bivariate level, baseline pupil size correlated with fluid intelligence in all the lighting conditions but with working memory capacity in only one of the conditions. The hierarchical linear models painted a similar picture but with the caveat that baseline pupil size did not correlate with fluid intelligence in the two brightest conditions. The structural equation models that included age as a correlated predictor also demonstrated that the baseline pupil size – working memory capacity is not as robust to the confound of age differences.

Given that pupil size is associated with changes in arousal we also measured subjective reports of arousal after each baseline condition. We found no within-subject changes in arousal associated with pupil size across the baseline conditions (Table S9 and Fig. S5). We did find between-subject differences such that, subjects that reported overall higher arousal had a larger baseline pupil size (Table S10). However, subjective reports of arousal did not correlate with any of the cognitive abilities.

4. General discussion

The basic question motivating this research was – do individual differences in baseline pupil size correlate with cognitive abilities? We believe the answer to that question is – Yes. In two studies we attempted to demonstrate that small and non-significant correlations between baseline pupil size and cognitive abilities (namely working memory capacity) reported by other researchers can be explained by a reduced mean and variance of baseline pupil size values due to too bright lighting conditions. In Study 1, we found that using a white background color on the monitor reduced the mean and variance of baseline pupil size values as well as the correlation with fluid intelligence, working memory capacity, and attention control. In Study 2, we showed that it is specifically a combination of the background color and brightness/contrast settings of the monitor that leads to a restriction of range on baseline pupil size values and a smaller correlation with cognitive abilities. Overall, we found that the baseline pupil size – working memory capacity relationship, compared with fluid intelligence, was less robust to restriction of range on pupil size and a small sample size.

In a recent meta-analysis Unsworth et al. (2020) reported a meta-analytic correlation between baseline pupil size and working memory capacity of $r = 0.01$. (Heitz et al., 2008; Tsukahara et al., 2016). To illustrate the extensive problem of reduced variance on pupil size, we recreated the meta-analytic table in Unsworth et al. (2020) and ordered the studies based on the standard deviation (inter-individual variability) of baseline pupil size in order of largest to smallest (Table S11). Critically, there are a large number of studies with very small mean and standard deviation values on pupil size. At worst, Unsworth and Robison (2017b) are near the minimal physiological limit (~2 mm) of pupil size values with a mean of 2.59 mm and standard deviation of 0.28.

In our reanalysis of Tsukahara et al. (2016) and current Studies 1 and 2, we found that standard deviations on baseline pupil size below 0.60 had smaller and non-significant correlations with cognitive abilities. Therefore, it would appear that standard deviations on baseline pupil size below ~0.60 mm are possibly problematic for individual difference research due to restriction of range (this should not be overly interpreted as a strict cutoff but only a very rough reference point). Given that ~67% (18 out of 27) of the studies had standard deviations below 0.60, it is hard to draw any strong conclusions from the meta-analysis (Unsworth et al., 2020). It should be noted, however, that there are correlations from studies above the 0.60 standard deviation cutoff that are also small and non-significant. The larger point is simply that, so far, there are very few studies that not only have a sufficient inter-individual variability in pupil size values but also use multiple measures of cognitive ability and recruit from a diverse and representative sample of sufficient size to detect the small correlation between pupil size and cognitive ability.

4.1. Methodological implications

Our results have important implications for researchers studying individual differences in both task-free and task-evoked changes in pupil size. Overall, individual differences in pupil size related cognitive abilities are not large; and therefore, careful consideration needs to be given to how pupil size and cognitive abilities are being measured. Otherwise, this could result in mixed findings simply due to methodological, and not theoretical, problems. We believe there are serious methodological issues across the majority of studies resulting in small and non-significant correlations between baseline pupil size and cognitive ability. Some of those issues include small sample size, measuring a cognitive ability with a single task, and recruiting from a homogenous and ability restricted sample. However, the primary reason we believe a large number of studies have found no relationship is due to a reduced inter-individual variance on baseline pupil size.

The largest impact on the variance and distribution of pupil size values is lighting conditions. Ideally, as researchers we would like to report objective values of lighting conditions that correspond most directly to the effect of luminance on the pupil. Based on our results from Study 2, we believe that the luminance levels reported by a light meter do not directly correspond to the luminance levels that will be picked up by the human eye. This is likely due to the light meter picking up a more even distribution of light from a wider field-of-view.

By examining the lux values in Table 8 and the mean and distribution of pupil values in Fig. 2 it becomes clear that the “Screen Lighting” measurement corresponds most closely to the changes in pupil size due to lighting conditions. The Background Color x Monitor Setting interaction seen in Fig. 2 matches closely to the “Screen Lighting” lux values. That is, the largest impact of lighting conditions on the distribution of pupil size values was the combination of a white background and bright monitor. These two conditions had “Screen Lighting” lux values that were much larger than any of the other conditions (208 vs. 1–15).

Although the “Participant View” may intuitively seem like the most direct way to measure the effect of luminance on the participant’s pupil, under certain circumstances it can actually be misleading. If we compare the “Participant View” lux values and observed pupil size values, it becomes more obvious how they do no match up. In three of the conditions with room lights on, the lux values are brighter (43, 44, and 47 lx) than the room lights off – white background – bright monitor condition (27 lx), see Table 8. Because the three conditions are brighter (according to the “Participant View” lux values) we should expect a smaller pupil size as well. That is not at all the case. Pupil size and inter-individual variability is actually quite a bit larger in those three “brighter” conditions (*M*(*SD*); 5.25 (0.84), 4.95 (0.80), and 4.51 (0.80) mm) compared to the other condition (M(SD); 3.07 (0.34) mm), see Table 10 and Fig. 2. This suggests that the reported lux values for “Participant View” do not correspond to the actual effect of luminance on pupil size. The “Room Lighting” lux values are informative as there is a general increase in pupil size values for room lights off compared to room lights on. Therefore, the advice to pupillometry researchers is to report lux values for the overall room lighting and screen lighting.

Finally, the black background conditions were less susceptible to differences in monitor settings than white backgrounds (and we would assume gray backgrounds as well, though to a lesser extent than white). Therefore, it might be advisable for pupillometry researchers to use a black background as a simple way to avoid potential differences in screen lighting across studies and research labs. However, using a black background with the room lights off might actually produce a ceiling effect on pupil size values. The black background with room lights off conditions, as seen in Fig. 2, had more negatively skewed distributions than the black background with room lights on. The conditions with a white background and dim monitor produced a wide and non-skewed distribution of pupil size values. Therefore, a gray background color may also be advisable. In general, it may be a good idea to pilot test what lighting conditions are required to obtain an appropriate range of baseline pupil size values.

Here is a list of general recommendations for any researchers investigating individual differences in either task-free or task-evoked pupil size:

- Recruit from as diverse and representative sample pool as possible. Do NOT use just university students (see Supplemental Materials).
- Use multiple tasks to measure a psychological construct like working memory capacity, fluid intelligence, or attention control.
- Make sure you are obtaining enough between-subject variability in baseline pupil size and other individual difference measures.
- Report lighting conditions and luminance values from at least two sources of illumination: overall room lighting and screen lighting (by placing the light meter directly on the monitor).
- Be very explicit about how luminance values were obtained. Did you use a light meter? How was it positioned?
- If possible, convert arbitrary pixel values to millimeters to allow for comparison across studies.
- Ask about caffeine and drug use, amount of sleep, arousal, age and any other factors that may have an impact on baseline pupil size.

4.2. Limitations and outstanding issues

The results from Study 1 were rather straightforward, and the first time our lab has observed small and non-significant correlations between baseline pupil size and cognitive ability. This was entirely due to measuring baseline pupil size in brighter lighting conditions. While the results from Study 2 overall strongly support the same conclusion, there are a few limitations that should be noted.

First of all, the sample size is considerably smaller than our previous studies which we believe resulted in rather small and non-significant correlations with working memory capacity across the board. This made the lighting manipulations less informative as to the interaction with working memory capacity. Nevertheless, it does suggest that the baseline pupil size – working memory capacity relationship is smaller and less robust overall. The correlations with fluid intelligence were also smaller than we had found previously and smaller than in Study 1, which again may be due to the smaller sample size.

Secondly, the bivariate correlations and simple slopes analysis from the hierarchical linear models have a slightly different interpretation for one of the lighting conditions; room lights off, white background, and bright monitor. This condition had a significant correlation with fluid intelligence and was relatively larger compared to some of the other darker conditions. However, the hierarchical linear model results showed only a significant Background Color x Monitor Setting interaction such that the slope was smaller and non-significant in the room lights off, white background, and bright monitor condition compared to the darker ones. Additionally, in the bivariate correlations, working memory capacity only correlated with baseline pupil size in the brightest lighting condition.

In a recent study, Unsworth et al. (2020) measured task-free baseline pupil size using a black background (with a white fixation cross) in a dark room and still found a small and non-significant correlation with working memory capacity, $r = 0.05$. They also had a large sample size, $n = 328$, and used multiple measures to get a composite factor on working memory capacity. Even though their mean baseline pupil size value was larger ($M = 4.89$) than their previous studies, the variance ($SD = 0.54$) was still problematic (smaller than 0.60). Their mean value was close to the mean values in some of the darker conditions in our studies, but their standard deviation value (0.54) was more similar to our brighter conditions.

Because they used a black background color in a dark room it cannot be claimed that they used too bright of lighting conditions. The only other major difference we can think of that might contribute to the reduced variance on pupil size is the demographics of the samples. In all our studies we recruited from a diverse sample of college students and non-college individuals in the Atlanta community. Whereas in their studies (Unsworth et al., 2019; Unsworth & Robison, 2015, 2017b), they recruited from a less diverse sample of primarily 18 and 19-year old’s at the University of Oregon. Regardless, their findings still suggest that the baseline pupil size –working memory capacity relationship is not robust to these various differences in reduced variance and sample demographics.

In fact, in supplemental analyses from Study 1 we found that if we only included the Georgia Tech students (convenience sampling from the local university where the study was conducted) then neither working memory capacity nor fluid intelligence correlated with baseline pupil size; however, for the non-Georgia Tech population (non-convenience sampling and more diverse demographics) both working memory capacity and fluid intelligence correlated with baseline pupil size (even after controlling for age). This pattern held up even after including students from Georgia State University (a neighboring college only miles away but more diverse on abilities) in the Georgia Tech sample. Further, these results were not explained by a reduced variability on baseline pupil size values because both samples had adequate variability. In summary, for university-only samples (from our two largest pools of university students) we do not find a correlation between baseline pupil size and cognitive ability, but we do find one in a more diverse and representative sample (mix of university and community subjects). Therefore, these supplemental findings highlight the importance of recruiting from as diverse a sample as possible for individual differences research on cognitive abilities and pupil size.

4.3. Theoretical considerations

It is now widely accepted that the size of the pupil of the eye can be used as an indicator of activity in the locus coeruleus (Joshi et al., 2016; Bruno Laeng & Sirois, 2012; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Rajkowski et al., 1993). We believe that the baseline pupil size – cognitive ability relationship is related to the functioning of the locus coeruleus-norepinephrine system. Research has implicated the locus coeruleus-norepinephrine system in attention and working memory (Aston-Jones & Cohen, 2005; Berridge & Waterhouse, 2003), in modulating global functional connectivity of brain networks (Berridge & Waterhouse, 2003; Guedj et al., 2016; Moore & Bloom, 1979; R. L. van den Brink, Pfeffer, & Donner, 2019; Warren et al., 2016), and with activity in default-mode brain regions (Murphy et al., 2014; Yellin, Berkovich-Ohana, & Malach, 2015). Additionally, greater strength of functional connectivity within brain networks such as default-mode and executive attention networks correlate with higher intelligence and working memory capacity (Finn et al., 2015; Gordon, Breeden, Bean, & Vaidya, 2012; Hellyer et al., 2014; Keller et al., 2015; Reineberg, Andrews-Hanna, Depue, Friedman, & Banich, 2015; Schultz & Cole, 2016; Smith et al., 2015; Song et al., 2009; Stevens, Tappon, Garg, & Fair, 2012; van den Heuvel, Stam, Kahn, & Hulshoff Pol, 2009).

Based on this evidence and our finding that baseline pupil size correlates with fluid intelligence, we proposed that *fluid intelligence is related to the functional organization of the resting-state brain arising from neuromodulatory role of the locus coeruleus-norepinephrine system* (Tsukahara et al., 2016). Specifically, larger baseline pupil size may indicate stronger functional connectivity in default-mode and executive attention networks, arising from optimal levels of tonic (baseline) locus coeruleus activity.

One possibility is that high ability individuals, even during a passive baseline state, are in a more task-ready state and at optimal levels of phasic locus coeruleus activity. This is supported by evidence that high ability individuals display less and more efficient reconfiguration of functional connectivity from resting-state to task engagement (Schultz & Cole, 2016).

Another possibility, is that high ability individuals more optimally regulate activity in the locus coeruleus (Unsworth & Robison, 2017a) and this allows them greater ability to switch from one mental state to another; specifically, for switching between a high tonic exploration mode (Aston-Jones & Cohen, 2005; Bornemann et al., 2010; van der Meer et al., 2010) to a phasic exploitation mode of locus coeruleus activity. This function of switching from one mental state to another would also be consistent with the network-reset theory of locus coeruleus function (Bouret & Sara, 2005) and various interpretations of the pupil size – cognitive ability relationship (Unsworth & Robison, 2017a; van der Meer et al., 2010).

5. Conclusion

There is a large body of literature suggesting that the locus coeruleus-norepinephrine system is essential to understanding the biological basis of higher-order cognitive abilities such as fluid intelligence and working memory capacity. The use of eye tracking technology to measure pupil size (known as pupillometry) has provided an accessible and non-invasive method for cognitive psychologists to study the locus coeruleus-norepinephrine system in relation to cognition (Joshi & Gold, 2019).

Although much work has been done to improve the methodology, pre-processing, and statistical analysis in pupillometry research there is still much room for improvement (Mathôt, Fabius, Van Heusden, & Van der Stigchel, 2018). The issue of different lighting conditions has recently been investigated by some researchers (Baldock, Kapadia, van Steenbrugge, & McCarley, 2019; Reilly, Kelly, Kim, Jett, & Zuckerman, 2019). However, not enough research has been done on how the distribution of pupil size values in too bright lighting conditions (restricting the range down to minimal physiological limits) and too dark lighting conditions (restricting the range up to maximal physiological limits) might bias results for both experimental and differential research on baseline and task-evoked changes in pupil size.

We have shown that, at least for differential research on baseline pupil size, that too bright lighting conditions will bias results towards the null due to restriction of range on pupil size values. Critically, there is no standard practice for objectively measuring lighting conditions that correspond most directly to the effect of luminance on pupil size. This can make comparisons between studies with mixed findings difficult to interpret. We have provided our recommendations on establishing standard practices for reporting lighting conditions.

Author note

This work was supported by Office of Naval Research Grant N00014-12-1-1011 to Randall W. Engle.

Data, analysis scripts, and results outputs are available at: https://osf.io/ajm4d/

Declaration of Competing Interest

None.

Appendix A. Reliability estimates for pupil measures in Study 2

Lighting condition	Baseline pupil size	Baseline pupil variability
Off_Black_Dim	0.97	0.54
Off_Black_Bright	0.97	0.37
On_Black_Dim	0.98	0.57

(*continued on next page*)

(*continued*)

Lighting condition	Baseline pupil size	Baseline pupil variability
On_Black_Bright	0.97	0.74
On_White_Dim	0.97	0.67
Off_White_Dim	0.97	0.64
On_White_Bright	0.83	0.75
Off_White_Bright	0.74	0.83

Note. Reliability was calculated as the correlation between the first and second half of the baseline period corrected with the spearman-brown prophecy formula. The split-half reliability of arousal ratings across the eight lighting conditions was 0.94.

Appendix B. Correlation table for SEMs in Study 2

	1.	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*
1. Gf	–	0.54	0.53	−0.28	0.17	0.21	0.20	0.16	0.26	0.17	0.25	0.24
2. WMC	0.54	–	0.43	−0.29	0.13	0.12	0.10	0.05	0.10	0.07	0.18	0.13
3. AC	0.53	0.43	–	−0.15	0.18	0.17	0.18	0.20	0.19	0.23	0.20	0.21
4. Age	−0.28	−0.29	−0.15	–	−0.31	−0.37	−0.38	−0.24	−0.25	−0.26	−0.26	−0.34
5. On_White_Bright	0.17	0.13	0.18	−0.31	–	0.78	0.70	0.57	0.60	0.60	0.73	0.75
6. On_White_Dim	0.21	0.12	0.17	−0.37	0.78	–	0.86	0.59	0.65	0.61	0.56	0.89
7. On_Black_Dim	0.20	0.10	0.18	−0.38	0.70	0.86	–	0.65	0.61	0.63	0.54	0.92
8. Off_Black_Dim	0.16	0.05	0.20	−0.24	0.57	0.59	0.65	–	0.75	0.87	0.57	0.69
9. Off_White_Dim	0.26	0.10	0.19	−0.25	0.60	0.65	0.61	0.75	–	0.80	0.76	0.67
10. Off_Black_Bright	0.17	0.07	0.23	−0.26	0.60	0.61	0.63	0.87	0.80	–	0.60	0.67
11. Off_White_Bright	0.25	0.18	0.20	−0.26	0.73	0.56	0.54	0.57	0.76	0.60	–	0.56
12. On_Black_Bright	0.24	0.13	0.21	−0.34	0.75	0.89	0.92	0.69	0.67	0.67	0.56	–

Computed correlation used pearson-method with pairwise-deletion.

Appendix C. Supplementary materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104643.

References

Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. In *, Vol. 80. Intelligence*. Elsevier Ltd. https://doi.org/10.1016/j.intell.2020.101440.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289–1292.

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*, 29–37.

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. https://doi.org/10.12688/wellcomeopenres.15191.1.

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 Review of Neuroscience, 28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709.

Baldock, J., Kapadia, S., van Steenbrugge, W., & McCarley, J. (2019). The effects of light level on the task-evoked pupil response during effortful listening. In *Poster presented at the meeting of Australasian Cognitive Neuroscience Society*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

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.

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

Bornemann, B., Foth, M., & Horn, J. (2010). *Mathematical cognition : Individual differences in resource allocation* (pp. 555–567). https://doi.org/10.1007/s11858-010-0253-x.

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences, 28*(11), 574–582. https://doi.org/10.1016/j.tins.2005.09.002.

van den Brink, R. L., Pfeffer, T., & Donner, T. H. (2019). Brainstem modulation of large-scale intrinsic cortical activity correlations. *Frontiers in Human Neuroscience, 13* (October), 1–18. https://doi.org/10.3389/fnhum.2019.00340.

Brown, S. M., Khanani, A. M., & Xu, K. T. (2004). Day to day variability of the dark-adapted pupil diameter. *Journal of Cataract and Refractive Surgery, 30*, 639–644. https://doi.org/10.1016/j.jcrs.2003.09.060.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*(5), 769–786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers, 17*(6), 652–655. https://doi.org/10.3758/BF03200977.

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment, 30*(1), 116–129. https://doi.org/10.1037/pas0000444.

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science, 11*(1), 133–155. https://doi.org/10.1177/1745691615596990.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). 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. https://psyarxiv.com/q985d/.

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. In *, Vol. 102. Princeton NJ educational testing service*. https://doi.org/10.1073/pnas.0506897102.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*(2), 190–193. https://doi.org/10.1177/1745691617720478.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *Vol. 44. The psychology of learning and motivation* (pp. 145–199). Elsevier Science.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Todd Constable, R. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience.*. https://doi.org/10.1038/nn.4135.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252.

Gordon, E. M., Breeden, A. L., Bean, S. E., & Vaidya, C. J. (2012). Working memory-related changes in functional connectivity persist beyond task disengagement. *Human Brain Mapping*. https://doi.org/10.1002/hbm.22230.

de Groot, S. G., & Gebhard, J. W. (1952). Pupil size as determined by adapting luminance. *Journal of the Optical Society of America, 42*(7), 492–495. https://doi.org/10.1364/josa.42.000492.

Guedj, C., Monfardini, E., Reynaud, A. J., Farnè, A., Meunier, M., & Hadj-Bouziane, F. (2016). Boosting norepinephrine transmission triggers flexible reconfiguration of brain networks at rest. *Cerebral Cortex*, 1–10. https://doi.org/10.1093/cercor/bhw262.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18*, 1279–1296.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.

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.

Hellyer, P. J., Shanahan, M., Scott, G., Wise, R. J. S., Sharp, D. J., & Leech, R. (2014). The control of global brain dynamics: Opposing actions of frontoparietal control and default mode networks on attention. *Journal of Neuroscience, 34*(2), 451–461. https://doi.org/10.1523/JNEUROSCI.1853-13.2014.

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.

van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *The Journal of Neuroscience, 29*(23), 7619–7624. https://doi.org/10.1523/JNEUROSCI.1443-09.2009.

Hughes, J. (2020). reghelper: Helper functions for regression analysis. https://cran.r-project.org/package=reghelper.

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., & Gold, J. I. (2019). *Pupil size as a window on neural substrates of cognition*. https://doi.org/10.31234/osf.io/dvsme.

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* (3756), 1583–1585.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66.

Keller, J. B., Hedden, T., Thompson, T. W., Anteraper, S. A., Gabrieli, J. D. E., & Whitfield-Gabrieli, S. (2015). Resting-state anticorrelations between medial and lateral prefrontal cortex: Association with working memory, aging, and individual differences. *Cortex, 64*, 271–280. https://doi.org/10.1016/j.cortex.2014.12.001.

Kovacs, K., & Conway, A. R. A. (2020). Process overlap theory, executive functions, and the interpretation of cognitive test scores: Reply to commentaries. *Journal of Applied Research in Memory and Cognition, 9*, 419–424. https://doi.org/10.1016/j.jarmac.2020.04.005.

Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science, 7*(1), 18–27. https://doi.org/10.1177/1745691611427305.

Laeng, B., & Sirois, S. (2012). *Perspectives on psychological science pupillometry: A window to*. https://doi.org/10.1177/1745691611427305.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–284. https://doi.org/10.1038/36846.

Lüdecke, D. (2020). sjPlot: Data visualization for statistics in social science. https://cran.r-project.org/package=sjPlot.

Martin, J. D., Tsukahara, J. S., Mashburn, C. A., Draheim, C., Shipstead, Z., Vogel, E. K., & Engle, R. W. (2019). The visual arrays task: Visual storage capacity or attention control?. Preprint on PsyArXiv https://psyarxiv.com/u92cm/.

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

van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., … Wartenburger, I. (2010). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology, 47*, 158–169. https://doi.org/10.1111/j.1469-8986.2009.00884.x.

Moore, R. Y., & Bloom, F. E. (1979). Central catecholamine neuron systems: Anatomy and physiology of the norepinephrine and epinephrine systems. *Annual Review of Neuroscience, 2*, 113–168. https://doi.org/10.1146/annurev.ne.02.030179.000553.

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping, 35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connll, R. G. (2011). Pupillometry and P3 index the locus coeruleus – Noradrenergic arousal function in humans. *Psychophysiology*, 1–12. https://doi.org/10.1111/j.1469-8986.2011.01226.x.

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002.

Psychology Software Tools, I. [E-P. 3. 0]. (2016). *E-Prime 3.0*. Psychology Software Tools, Inc. https://www.pstnet.com.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (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*, 974.

Raven, J., Raven, J. C., & Court, J. (1998). Manual for Raven's progressive matrices and vocabulary scales. In *Raven manual*.

Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods, 51*, 865–878. https://doi.org/10.3758/s13428-018-1134-4.

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage, 104*, 69–78. https://doi.org/10.1016/j.neuroimage.2014.09.045.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin and Review, 26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y.

Ruud Van Den Brink, L., Murphy, P. R., & Nieuwenhuis, S. (2016). *Pupil diameter tracks lapses of attention* (pp. 1–16). https://doi.org/10.5061/dryad.mp332.

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

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Schultz, D. H., & Cole, M. W. (2016). Higher intelligence is associated with less task-related brain network reconfiguration. *Journal of Neuroscience, 36*(33), 8551–8561. https://doi.org/10.1523/JNEUROSCI.0358-16.2016.

Shipstead, Z., Lindsey, D. R. B., 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.

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E., Glasser, M. F., … Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*. https://doi.org/10.1038/nn.4125.

Song, M., Liu, Y., Zhou, Y., Wang, K., Yu, C., & Jiang, T. (2009). Default network and intelligence difference. *IEEE Transactions on Autonomous Mental Development, 1*(2), 2212–2215. https://doi.org/10.1109/IEMBS.2009.5334874.

Stawarczyk, D., & D'Argembeau, A. (2016). Conjoint influence of mind-wandering and sleepiness on task performance. *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), 1587–1600. https://doi.org/10.1037/xhp0000254.

Stevens, A. A., Tappon, S. C., Garg, A., & Fair, D. A. (2012). Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PLoS One, 7*(1), 1–10. https://doi.org/10.1371/journal.pone.0030468.

Thurstone, L. L. (1938). *Primary mental abilities*.

Tsukahara, J. S. (2019). *pupillometry: An R package to preprocess pupil data*. https://doi.org/10.5281/zenodo.4464854.

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.

Ullwer, D., Bornemann, B., Ries, J., Horn, J., Foth, M., Preusse, F., Wartenburger, I., & van der Meer, E. (2003). *Resource allocation and problem solving strategies during a geometric analogy task in individuals differing in fluid intelligence*.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/BF03192720.

Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Is working memory capacity related to baseline pupil diameter? *Psychonomic Bulletin & Review*. 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. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-016-0417-4.

Unsworth, N., & Robison, M. K. (2017a). A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review, 1957*. 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*(12), 1962–1987. https://doi.org/10.1037/xlm0000421.

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*, 1–20. https://doi.org/10.3758/s13415-019-00709-z.

Warren, C. M., Eldar, E., van den Brink, R. L., Tona, K.-D., van der Wee, N. J., Giltay, E. J., … Nieuwenhuis, S. (2016). Catecholamine-mediated increases in gain enhance the precision of cortical representations. *Journal of Neuroscience, 36*(21), 5699–5708.

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

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.

Yellin, D., Berkovich-Ohana, A., & Malach, R. (2015). Coupling between pupil fluctuations and resting-state fMRI uncovers a slow build-up of antagonistic responses in the human cortex. *NeuroImage, 106*, 414–427. https://doi.org/10.1016/j.neuroimage.2014.11.034.