The relationship between baseline pupil size and intelligence

Jason S. Tsukahara *, Tyler L. Harrison, Randall W. Engle

Georgia Institute of Technology, United States

ARTICLE INFO

ABSTRACT

Pupil dilations of the eye are known to correspond to central cognitive processes. However, the relationship between pupil size and individual differences in cognitive ability is not as well studied. A peculiar finding that has cropped up in this research is that those high on cognitive ability have a larger pupil size, even during a passive baseline condition. Yet these findings were incidental and lacked a clear explanation. Therefore, in the present series of studies we systematically investigated whether pupil size during a passive baseline is associated with individual differences in working memory capacity and fluid intelligence. Across three studies we consistently found that baseline pupil size is, in fact, related to cognitive ability. We showed that this relationship could not be explained by differences in mental effort, and that the effect of working memory capacity and fluid intelligence on pupil size persisted even after 23 sessions and taking into account the effect of novelty or familiarity with the environment. We also accounted for potential confounding variables such as; age, ethnicity, and drug substances. Lastly, we found that it is fluid intelligence, more so than working memory capacity, which is related to baseline pupil size. In order to provide an explanation and suggestions for future research, we also consider our findings in the context of the underlying neural mechanisms involved.

1. Introduction

Starting in the 1960s it became apparent to psychologists that the size of the pupil is related to more than just the amount of light entering the eyes. Pupil size also reflects internal mental processes. For instance, in a simple memory span task, pupil size precisely tracks changes in memory load, dilating with each new item held in memory and constricting as each item is subsequently recalled (Hess & Polt, 1964; Kahneman & Beatty, 1966). This research established the use of pupil dilations as an indicator of momentary changes in arousal, mental effort, and attention (Beatty & Lucero-Wagoner, 2000; Hess & Polt, 1960).

Because pupil dilations occur for a wide variety of tasks involving mental effort, psychologists had inferred that the task-evoked pupillary response was reflective of central brain processes (Beatty, 1982). For some, this was seen as providing an opportune way to study the dynamics of cognitive brain function (Beatty & Lucero-Wagoner, 2000). Until more recently, though, the method of measuring pupil size to study brain function did not gain much traction in the field. It was suspected that the reason for this was, "pupillometry is not widely employed in cognitive psychophysiology because the pupil lacks face validity as a measure of brain function" (Beatty & Lucero-Wagoner, 2000).

---

* Corresponding author at: School of Psychology, Georgia Institute of Technology, 654 Cherry St., Atlanta, GA 30332, United States.
E-mail address: jtsukahara3@gatech.edu (J.S. Tsukahara).

The discovery that changes in pupil size correspond to activity in the locus coeruleus was pivotal in establishing pupil size as an important indicator of brain function. (Joshi, Li, Kalwani, & Gold, 2016; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Rajkowski, Kubiak, & Aston-Jones, 1993; Varazzani, San-Galli, Gilardeau, & Bouret, 2015). The locus coeruleus is a region in the brain stem that has projections throughout the brain and is the main source of norepinephrine in the central nervous system (Moore & Bloom, 1979). The method of measuring pupil size is becoming more common in the fields of psychology and neuroscience today due to a better understanding of the brain regions associated with pupil size, along with theoretical advancements in the role these regions play in central cognition (Aston-Jones & Cohen, 2005b).

While most researchers have focused on within-individual changes in pupil size in relation to cognition and brain activity, the present study investigates differences in baseline pupil size **between** individuals of differing cognitive ability. This investigation was motivated by preliminary findings in our lab that larger baseline pupil size was associated with better performance on the operation span task, a measure of working memory capacity (Heitz, Schrock, Payne, & Engle, 2008). That is, high span subjects had larger pupils than low span subjects even during a "passive" baseline (in the absence of performing any specific cognitive task). However, the relationship was incidental, a number of potential confounds were present, and it was only treated tangentially. Therefore, this finding required further verification.

Besides this being a fascinating relationship though, it is important to consider this in the context of how it might relate to underlying differences in brain function. It turns out that, the prefrontal cortex, one of the crucial areas related to higher order cognitive abilities such as working memory capacity and fluid intelligence (Jung & Haier, 2007; Kane & Engle, 2002), receives a large amount of projections from the locus coeruleus. Furthermore, the locus coeruleus-norepinephrine system is recognized as playing a central role in the cognitive control of behavior through its neuromodulatory effects on regions such as the prefrontal cortex (Aston-Jones & Cohen, 2005a). In light of these advancements, we feel that if the relationship of cognitive ability to pupil size is real it may point to important dynamics in brain function that give rise to individual differences in cognitive ability.

In the present series of studies, we investigated whether the relationship between baseline pupil size and cognitive ability is real, or an artifact, by controlling for a number of potential confounding variables. At the time of the Heitz et al. (2008) paper, we were not sure whether this relationship was real or due to various confounding factors; such as age, mental effort, or experience in the lab. Although, one other lab has shown that performance on the Ravens advanced progressive matrices, a measure of fluid intelligence, was similarly associated with baseline pupil size (van der Meer et al., 2010). Nevertheless, for both studies the relationship was an incidental finding, was treated tangentially, and lacked a clear explanation. Therefore, the present study is the first to systematically investigate the relationship between baseline pupil size and cognitive abilities. Furthermore, in the discussion, we provide an explanation of this finding in terms of the potential underlying neural mechanisms.

It is important to note that Heitz et al. (2008) and van der Meer et al. (2010) used only a single task as their measure of working memory capacity and fluid intelligence, respectively. It is problematic to make inferences about the relationship of a variable with constructs such as working memory capacity or fluid intelligence using a single measure because all measures are multiply determined. For example, the Raven Matrices task is the most commonly used measure of fluid intelligence. However most large scale studies show that only 50–60% of the variance in Raven scores is attributable to fluid intelligence (Carroll, 1993; Jensen, 1998). The remaining 40–50% of the variance in Raven scores can be attributed to myriad other factors such as spatial rotation skills, experience with matrix problems and motivation. Measuring a construct with a single task not only measures the abilities of interest but also other unrelated and possibly unknown factors that are specific to that particular task. Thus, the only way to establish that a variable such as pupil size is associated with a general construct, such as working memory capacity or fluid intelligence, is to use multiple indicators from a range of domains.

Therefore, in all studies we used multiple measures of working memory capacity and/or fluid intelligence, allowing us to make claims at the construct level. At the construct level we refer to working memory capacity as the ability to actively maintain attention on information in the face of interference and is associated with performance across a wide range of cognitive domains (Ackerman, Beier, & Boyle, 2005; Conway, Kane, & Engle, 2003; Harrison, Shipstead, & Engle, 2014; Heitz et al., 2006; Shipstead, Lindsey, Marshall, & Engle, 2014). Fluid intelligence refers to the ability to solve novel reasoning problems and is highly correlated with working memory capacity (Ackerman et al., 2005; Conway et al., 2003; Engle, Tuholski, Laughlin, & Conway, 1999; Harrison et al., 2014; Heitz et al., 2006; Kane et al., 2004; Shipstead et al., 2014).

Baseline pupil size was measured during a "passive" baseline while subjects stared at a fixation on a computer monitor. In Study 1, we tested whether mental effort can account for the relationship between baseline pupil size and cognitive ability. In Study 2, we assessed the reliability of baseline pupil size measures and controlled for the effects of familiarity of the environment on the relationship with cognitive ability. In Study 3, we investigated whether working memory capacity or fluid intelligence is uniquely related to baseline pupil size, while also controlling for a number of potential confounds not addressed in Studies 1 and 2.

2. Study 1: Mental effort

Given that change in pupil size is commonly used to assess the amount of mental effort in which one is allocating, it is important to test whether mental effort can account for the relationship between baseline pupil size and cognitive ability. It may be that there are differences in the amount of attention or effort high and low cognitive ability subjects allocate during a

"passive" baseline. Therefore, in Study 1, we tested this possibility by measuring baseline pupil size in the absence of an obvious task, and whether this difference persisted as cognitive load increased in an attention-demanding cognitive task.

2.1. Method

2.1.1. Subjects

Subjects were selected from a larger subject pool in which measures of working memory capacity (WMC) were already obtained. Forty subjects were recruited from the upper (high WMC; n = 20) and lower (low WMC; n = 20) quartile of a WMC composite score. Subjects were between the ages of 18–35 and had corrected-to-normal vision. All subjects signed a consent form approved by the IRB.

2.1.2. Stimuli and procedure
2.1.2.1. Working memory tasks. The composite score for high and low WMC subjects was an average of z-scores on three complex span tasks; the operation span, reading span, and symmetry span tasks. The complex span tasks consisted of alternating storage and processing sub-tasks (Unsworth, Heitz, Schrock, & Engle, 2005). Subjects first solved a processing task followed by the presentation of a single memory item. This alternation repeated until a range (set-size) of memory items had been presented. On the recall screen subjects had to recall the memory items in the correct order by clicking the appropriate item from a matrix on the computer. In all the complex span tasks, the dependent variable was the total number of items recalled in their correct serial position (partial score).
Operation span. The automated operation span task required subjects to remember a series of letters presented in alternation with simple math equations, which they were required to solve. Set-sizes ranged from 3 to 7 letters and each set occurred 3 times.
Symmetry span. The automated symmetry span required subjects to remember a series of spatial locations presented alternately with a pattern of blocks and they had to decide whether the pattern was symmetrical around the vertical midline. Set-sizes ranged from 2 to 5 spatial locations and each set occurred 3 times.
Reading span. The automated reading span required subjects to remember a series of letters presented at the end of a sentence. They judged whether the sentence made sense and then saw the letter. Set-sizes ranged from 3 to 7 letters and each set occurred 3 times.

2.1.2.2. Pupil measures. Pupil size was measured before beginning any experimental tasks using an Applied Science Laboratories (ASL) model 5000 eye-tracker unit, sampling at 60 Hz. Pupil data were recorded using software provided by ASL. Pupil size was recorded prior to receiving instructions for experimental tasks. After a 5-point calibration procedure, pupil size was recorded while subjects stared at a white fixation (font size: 28) against a black background for 2 min. Subjects were seated 60–80 cm from the monitor. No devices, such as a chin-rest, were used to stabilize the subject's head position. Illumination was controlled for as pupil size was measured in the same room for all subjects. Missing data due to blinks and subject's head movements were excluded from analysis. No subjects or trials were eliminated due to excessive loss of data. Baseline pupil size was calculated as the average pupil diameter (mm) during the fixation.

Subjects then performed a simple letter span task. Memory items were single letters presented visually and for a duration of 1 s with a 250 ms inter-stimulus interval. The number of memory items (set-size) presented on each trial was either 4, 6, or 8. Each set-size occurred four times and occurred in an ascending order. Subjects were instructed to remember as many letters as they could and in the order that they were presented. A delay of 30 s followed the last presented memory item, after which a recall screen appeared. Subjects used a computer mouse to select as many letters as they could recall and in the order they were presented. Feedback informed subjects as to how many memory items they correctly recalled in serial order out of the total set-size. Pupil diameter was averaged over the delay period as the measure of pupil size during the task.[1] For access to the data, send a request to the corresponding author (Jason Tsukahara: jtsukahara3@gatech.edu).

2.2. Results

Summarized in Fig. 1, we found that for the "passive" baseline high WMC subjects pupil diameters were nearly a millimeter larger than those with low WMC, a difference of about 0.97 mm, $t(38) = 3.48$, $p < 0.05$, $d = 1.10$. To put this into comparison, task-evoked pupil dilations are no larger than 0.5 mm and usually smaller (Beatty & Lucero-Wagoner, 2000). A difference of about 1 mm is very large and one that is likely observable to the naked eye, if the pupil is surrounded by a light-colored iris. The change in pupil diameter over levels of memory load, seen in Fig. 1, reflects the increase in mental effort, $F(3,114) = 11.93$, $p < 0.05$, $\eta^2_p = 0.24$. The important finding here was that pupil diameter increased as a function of memory load by the same amount for high and low WMC subjects (Load $\times$ WMC interaction, $F < 1$). High and low WMC pupil size differences remain even after they engaged in a memory task shown by the main effect of WMC, $F(1,38) = 13.54$,

---

[1] Typically, task-related changes in pupil size are calculated by subtracting out pre-trial baseline pupil size. This corrects for any differences at baseline that might exist. However, in this study we are interested in those baseline differences and if they persist over and above task-related changes in pupil size.
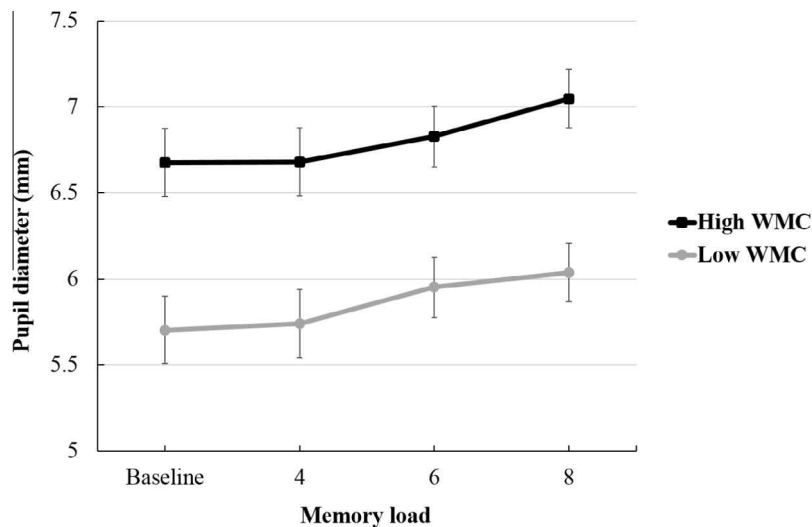
Fig. 1. Mean pupil diameter at baseline and set sizes of 4, 6 and 8 for high WMC (n = 20) and low WMC (n = 20). Error bars represent standard error of the mean.

$p < 0.05$, $\eta^2_p = 0.28$. Therefore, the relationship between pupil size and cognitive ability is a persistent one that remains even as subjects engaged in a demanding cognitive task.

Another possibility is that the high WMC subjects initially have higher levels of arousal, attention, or effort at the start of the baseline measure; and therefore, we are finding differences in baseline pupil size related to cognitive ability. To test this, we analyzed changes in pupil size across four 30-s blocks of the two-minute baseline measure, as shown in Fig. 2. We found that baseline pupil size did decrease over the two-minutes, $F(3,114) = 6.48$, $p < 0.05$, $\eta^2_p = 0.15$, suggesting higher levels of arousal at the start of the baseline measure. However, this decrease in pupil size over time was the same for both high and low WMC subjects (Time × WMC interaction, $F < 1$). Therefore, initial levels of arousal at the start of baseline measures could not account for the relationship between baseline pupil size and cognitive ability.

2.3. Discussion

We found that higher cognitive ability subjects had a larger pupil size in a "passive" baseline. Given that pupil size has commonly been used as an indicator of mental effort (for a review: Beatty & Lucero-Wagoner, 2000; Heitz et al., 2008) it was important to test if mental effort can account for this finding. Our results demonstrate that pupil size differences between high and low cognitive ability subjects remained as they engaged in a demanding cognitive task. We also showed that initial levels of arousal at the beginning of baseline measures could not account for this relationship.

Yet, this does not entirely rule out the effect of mental effort. For instance, it may be that differences in mental effort between high and low ability subjects remain as cognitive load increases; and therefore, we found no Load × WMC interaction. However, while this may be true, there are two reasons why we believe mental effort cannot completely account for our finding. (1) The difference in baseline pupil size that we found, 0.97 mm, is much larger than what is typically found for changes in pupil dilation due to effort, ∼0.5 mm, (Beatty & Lucero-Wagoner, 2000). (2) Fig. 1 nicely illustrates that high cognitive ability subject's pupil size at baseline is larger than low cognitive ability subject's pupil size even at the highest set-size. If our finding were due to differences in mental effort, then it would have to be reasoned that high ability subjects are giving more mental effort at a "passive" baseline than low ability subjects are while maintaining 8 items in memory. Therefore, mental effort cannot not explain differences in pupil size between high and low cognitive ability individuals.

3. Study 2: Familiarity with the environment and reliability

Next we address two potential issues. One is, what is the reliability of baseline pupil measures within the individual? Another is whether the difference may reflect a familiarity with the university setting rather than differences in cognitive ability. The reason for this concern is that our subjects come from both the university setting and the greater Atlanta community. It is possible that our high cognitive ability individuals are simply more familiar and more comfortable in a psychology lab at a large university. High cognitive ability subjects may have more computer experience, more likely to be a college student (vs. non-students from the community), or more familiar with testing environments. As a result, if familiarity of the environment affects arousal levels or allocation of attention to the environment this could possibly explain differences in pupil size at baseline. To address this issue we decided to get multiple measures of pupil size as subjects came into the lab for 23 sessions for a study on another topic.
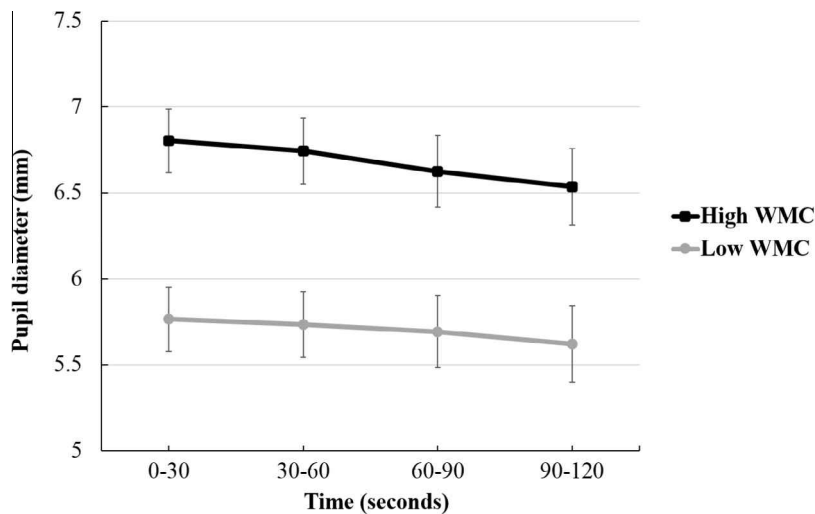
Fig. 2. Mean pupil diameter at 30-s intervals across 2 min of the baseline measure for high (n = 20) and low (n = 20) working memory capacity (WMC) subjects. Error bars represent standard error of the mean.

Baseline pupil size was measured at three different times; first session (Time 1), 12th session (Time 2), and 23rd session (Time 3). These three sessions occurred across a range from 3.5 weeks to 16.5 weeks. Baseline pupil size was recorded during a task-free rest period, and before subjects began any other experimental tasks for that day. In addition to working memory capacity, the relationship between pupil size and fluid intelligence was assessed.

3.1. Method

3.1.1. Subjects

We obtained baseline pupil measures from a sample of subjects participating in a cognitive training study, having nothing to do with our pupil measurements. For the training study, we recruited 114 subjects from the lower (n = 56) and upper (n = 58) third of a working memory capacity (WMC) composite score. Our decision for a tertiary split, compared to a quartile split in Study 1, was based on criteria for the cognitive training study and not baseline pupil measures. Subjects were between the ages of 18–35 and had corrected-to-normal vision. All subjects signed a consent form approved by the IRB. Due to technical issues with the eye-tracker, unable to calibrate eye-tracker, or excessive amounts of missing baseline pupil data, four subjects had missing data for one of the pupil measurements.

3.1.2. Stimuli and procedure
3.1.2.1. Working memory tasks.
The composite score for high and low WMC subjects was an average of z-scores on three complex span tasks; operation span, symmetry span, and rotation span (Unsworth et al., 2005). The operation span and symmetry span tasks were the same as in Study 1. For all complex span tasks, the dependent variable was the total number of items recalled in their correct serial position (partial score).

Rotation span. The automated rotation span task required subjects to remember a series of directional arrows (8 directions) of varying size (small or large) in alternation with a mental rotation task in which they had to mentally rotate and decide if a letter was mirror reversed or not. Set-sizes ranged from 2 to 5 memory items and each set occurred 3 times.

3.1.2.2. Fluid intelligence tasks.
The dependent variable for each fluid intelligence task was the number of correct responses. Scores on these tasks, from subjects who took part in pupillometry measures, were z-scored and averaged to form a composite on fluid intelligence (Gf).

Raven advanced progressive matrices (Raven, Raven, & Court, 1998). For each problem, subjects were presented with a 3 × 3 matrix containing 8 abstract figures and one empty box. Subjects were instructed to select of several response options which abstract figure belonged in the empty box to fit the pattern. Ten minutes were given to complete 18 of the odd number problems.

Letter sets (Ekstrom, French, Harman, & Dermen, 1976). For each problem, subjects were presented with five sets of four-letter sequences. They were instructed to indicate which set does not belong by discovering the rule that is common to only four of the sets. Five minutes were given to complete 30 problems.

Number series (Thurstone, 1938). For each problem, subjects were presented with a series of numbers. They were instructed to select which of several options completed the series. Five minutes were given to complete 15 problems.

3.1.2.3. Pupil measurement.
A SensoMotoric Instruments (SMI) RED 250 eye-tracking unit was used with a sampling rate of 60 Hz. Pupil data were recorded using iView X software provided by SMI. The SMI eye-tracking software uses a formula to

derive pupil diameter in millimeters from pixel values. The training study consisted of 23 days of training on a working memory capacity task. Pupil measurements were taken at the beginning of three general testing sessions, the 1st, 12th and 23rd sessions. For all three time points, pupillary measures were taken before beginning any tasks for that day.

Subjects were individually brought into a room for pupil measures. After a 5-point automatic calibration procedure, subjects stared at a gray fixation (font size: 30) against a black background for 21 s while their pupil diameters were recorded. Subjects were seated 60–80 cm from the monitor. No devices, such as a chin-rest, were used to stabilize the subject's head position. Pupil measurements for all subjects and all sessions were recorded in the same room to control for differences in illumination. Because baseline pupil measurements were taken for only 21 s, missing data due to blinks and subject's head movements were linearly interpolated. Subjects with excessive missing data were removed from analysis. Baseline pupil size was calculated as the average pupil diameter (mm) during the fixation. For access to the data, send a request to the corresponding author (Jason Tsukahara: jtsukahara3@gatech.edu).

3.2. Results

The change in baseline pupil size over the 23 sessions is summarized in Fig. 3. At Time 1, pupil diameter was on average 0.62 mm larger for the high WMC than low WMC subjects, $t(111) = 3.05$, $p < 0.05$, $d = 0.57$. There was also an effect of session, mean pupil diameter decreased over the three sessions, $F(2,216) = 5.19$, $p < 0.05$, $\eta^2_p = 0.046$. This suggests that pupil size decreased as subjects became more familiar with the environment. There was a main effect of WMC on pupil size, $F(1,108) = 6.97$, $p < 0.05$, $\eta^2_p = 0.06$. More importantly, the difference in pupil size between high and low WMC subjects remained over the three measurements ($F < 1$). Therefore, even though mean pupil diameter decreased over the three sessions the difference in pupil size between the two WMC groups remained significant.

This finding addresses a potential problem with the previous studies and we have now shown that familiarity of the environment does not account for the relationship between WMC and pupil size. Furthermore, the reliability of pupil size over time is high, as indicated by the high correlations amongst Times 1, 2 and 3 shown in bold in Table 1, ranging from 0.77 to 0.84.

Additional analyses were performed to determine if the relationship of Gf to pupil size remains unchanged over time and as subjects become familiar with the environment (van der Meer et al., 2010). As a first step, to simplify the analysis, pupil diameter was averaged across the three sessions to create an average pupil diameter variable. Average pupil diameter positively correlated with Gf, $r(102) = 0.37$, $p < 0.05$. To see if this correlation changed across the three sessions, Gf was correlated with pupil size at Time 1, Time 2 and Time 3. All correlations were significant at $p < 0.05$, and are shown in Table 1. Numerically the correlations decreased somewhat over the three sessions but this decrease was not significant (Time 1 ($r = 0.41$), Time 3 ($r = 0.30$), $z = 0.88$, $p > 0.05$). Therefore, the correlation between fluid intelligence and pupil size remained stable over the three measurements. This finding mirrors what we found for WMC. That is the relationship between Gf and pupil size does not change over time nor as subjects become more familiar with the environment.

3.3. Discussion

These findings extend those from Study 1 by showing that the effect of working memory capacity on pupil size does not change over time and as subjects become familiar with the environment. Importantly, we also showed that baseline pupil size is a reliable measure, a necessary requirement for psychophysiological variables. This strongly suggests that pupil size is a stable and reliable characteristic associated with individual differences in cognitive ability. Given that subjects were split on working memory capacity, this makes it difficult to assess the unique contributions of working memory and fluid intelligence to baseline pupil size. Although working memory capacity and fluid intelligence are highly correlated constructs, they are not the same (Ackerman et al., 2005; Conway et al., 2003; Engle et al., 1999; Harrison et al., 2014; Heitz et al., 2006; Kane et al., 2004; Shipstead et al., 2014); therefore, it is important to investigate their unique contributions to baseline pupil size.

4. Study 3: Working memory capacity, fluid intelligence, and baseline pupil size

The previous two studies established that baseline pupil size is a reliable physiological variable related to cognitive ability. However, one of the major limitations is the use of extreme groups design. It may be tempting to conclude from Study 1 and 2 that baseline pupil size relates to individual differences in working memory capacity, because high working memory individuals showed larger pupil size than low working memory individuals. Because working memory capacity and fluid intelligence are so highly correlated, when using an extreme groups design working memory capacity is confounded with fluid intelligence. Therefore, while we know baseline pupil size is related to individual differences in cognitive ability we do not know if it is uniquely related to either working memory capacity or fluid intelligence. In Study 3, we address the major limitation of Studies 1 and 2 by using the full scale sample allowing us to look at the full range of individual differences in working memory capacity and fluid intelligence and examine their unique relationships to baseline pupil size.

We also investigated how task-related pupil dilations over varying levels of cognitive load interacts with cognitive abilities. It is known that, in a simple memory span task, the pupils will dilate as each new item is held in memory
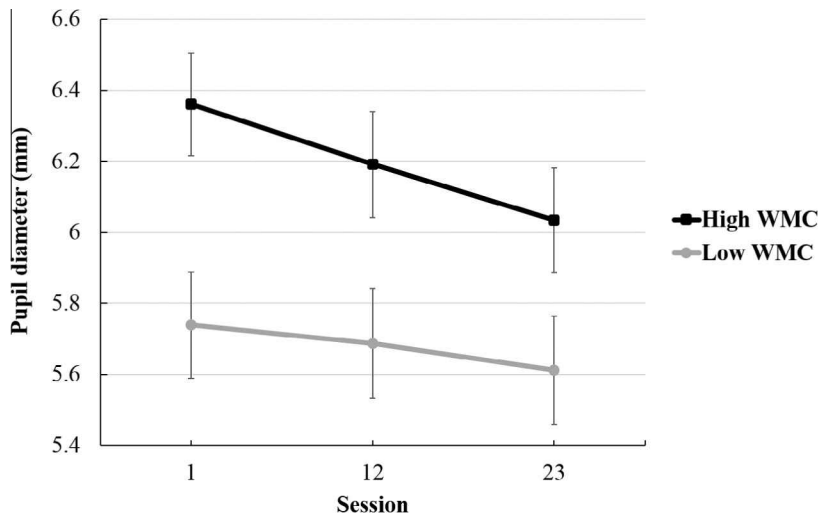
Fig. 3. Mean pupil diameter for high WMC (n = 57) and low WMC (n = 53) subjects at sessions 1, 12 and 23. Error bars represent standard error of the mean.

Table 1
Correlation matrix for fluid intelligence (Gf) and mean pupil diameter (mm) at Times 1, 2, and 3.

	Gf	Time 1	Time 2	Time 3
Gf	–			
Time 1	0.41	–		
Time 2	0.31	**0.84**	–	
Time 3	0.30	**0.77**	**0.77**	–

Note: All correlations are significant at $p < 0.05$. Time 1 = session 1; Time 2 = session 12; Time 3 = session 23; Gf = fluid intelligence.

(Granholm, Asarnow, Sarkin, & Dykes, 1996; Kahneman & Beatty, 1966; Peavler, 1974). However, there are inconsistent results as to what happens after memory capacity has been reached, assuming a short-term capacity of 7 ± 2 items (Miller, 1956). Peavler (1974) found that pupil diameter increased until it reached asymptote at about 8 items, remaining at peak diameter even at a load of 13 items. However, Granholm et al. (1996) found that pupil diameter peaked at about 9 items but then started to decline instead of maintaining asymptote. Research has suggested that low ability individuals show larger pupil dilations for easy problems and high ability individuals show larger pupil dilations for more difficult problems (Ahern & Beatty, 1979; Heitz et al., 2008; van der Meer et al., 2010). Therefore, we expect an interaction between memory load and cognitive ability on task-related pupil dilations.

4.1. Method

4.1.1. Subjects

A total of 358 subjects took part in four 2-h sessions in which they were tested on a wide-variety of cognitive tasks. No subject had participated in a study in our lab previously. Subjects were between the ages of 18–35 and had corrected-to-normal vision. Due to technical issues with the eye-tracker, unable to calibrate eye-tracker, or excessive amounts of missing baseline pupil data, the total number of subjects was reduced to 337.

4.1.2. Materials and procedure

Subjects participated in four sessions that lasted approximately 2 h in which they completed a battery of cognitive tasks. Included in this battery were the measures of working memory capacity (WMC) and fluid intelligence (Gf) described below. We measured baseline pupil size at the beginning of Session 4 before subjects started any tasks for that day. Immediately following baseline pupil measures subjects performed a simple memory-span task to measure task-evoked pupil dilations.

4.1.2.1. Working memory tasks. Measures of WMC consisted of the operation span, rotation span, and symmetry span tasks (Unsworth et al., 2005). These tasks were identical to those used in Study 2 except for the following. Operation span set-sizes ranged from 3 to 9 items and each set occurred 2 times. Symmetry and Rotation span tasks set-sizes ranged from 2 to 7 items and each set occurred 2 times. For all complex span tasks, the dependent variable was the total number of items recalled in their correct serial position (partial score).

Gf tasks.   Measures of Gf consisted of the Raven Advanced Progressive Matrices (Raven et al., 1998), Letter Sets (Ekstrom et al., 1976), and Number series (Thurstone, 1938) as described in Section 3.1.

Due to occasional issues with program execution, errors in saving data files, and subjects not completing a task, 13% of subjects had missing values for at least one task. Over 95% of subjects with missing values had no more than a single task missing from each construct. We therefore imputed missing values (Cohen, Cohen, West, & Aiken, 1983). Missing values on the WMC and Gf tasks were imputed with a maximum likelihood estimation using EQS 6.2. After imputation, scores on all tasks were z-scored and averaged to create a composite on WMC and Gf.

4.1.2.2. Demographics.   Given that the size of the pupil is affected by a variety of factors besides locus coeruleus activity, such as age and some drug substances, it is important to account for these. Nine different demographic variables were assessed: Ethnicity, Age (in years), College Student, Nicotine, Medications, Gender, Handedness, Caffeine, Alcohol, Sleep.

All demographics were self-reported. At the end of Session 4, the same day as pupil measurements, subjects were asked about: the amount of sleep they got the previous night, their use of nicotine (in the last 10 h), medications (that might affect their attention and memory, in the last 24 h), caffeine (in the last 8 h), and alcohol (more than two drinks in the last 24 h).

4.1.2.3. Pupil measures.   A SensoMotoric Instruments (SMI) RED 250 eye-tracking unit was used with a sampling rate of 60 Hz. Pupil data were recorded using iView X software provided by SMI. The SMI eye-tracking software uses a formula to derive pupil diameter in millimeters from pixel values.

Baseline pupil size.   Pre-experimental baseline pupil size was measured at the beginning of Session 4 before subjects began any tasks for that day and in a dimly lit room. All subject's pupils were measured in the same room to control for any differences in illumination. After a 5-point automatic calibration procedure, subjects passively viewed a gray fixation (font size: 30) against a black background on a computer monitor for 30 s. Subjects were seated 60–80 cm in front of the monitor. No devices, such as a chin-rest, were used to stabilize the subject's head position. Missing data due to blinks and subject's head movements were linearly interpolated. Subjects with excessive missing data were removed from analysis. Pre-experimental baseline pupil size was calculated as the average pupil diameter (mm) during fixation.

Task-related pupil size.   Task-related pupil changes were measured while subjects performed a simple memory span task. In this task, subjects were instructed to remember a series of visually presented letters and in the order they were presented. At recall they were presented with an array of 20 test letters and were instructed to click on each letter they could recall and in the order they were presented. Memory set-sizes of 3, 6, 9, or 12 occurred in a random order for a total of two times for each set-size. Although the order of occurrence was random, all subjects were given the same order of set-sizes.

A trial lasted from the presentation of the first memory item to the end of the recall phase and there was a 9 s intertrial interval. Stimuli were presented against a black background with a central fixation. Each memory item was presented for 750 ms. For half of the subjects the memory items were presented in cyan (blue) and for the other half lime (green) font color. The reason for this counterbalance was due to the addition of two other memory load conditions. In these conditions, there were both target items and distractor items, presented in either blue or green font. Font colors were counterbalanced between subjects. Subjects were instructed to simply ignore the distractor items and only remember the targets. In all analyses, these trials were collapsed into the non-distractor load trials as there was no difference in phasic pupil response between conditions. There was an interstimulus interval of 1000 ms and a delay of 2000 ms between the last memory item and the recall screen.

There were two measures of pupil size:[2] (1) Pretrial baseline pupil size was average pupil diameter (mm) during the 4 s before the start of a trial. (2) Pupil dilation was calculated as the difference between maximum pupil diameter observed after presentation of a memory item and pretrial baseline pupil size. For access to the data, send a request to the corresponding author (Jason Tsukahara: jtsukahara3@gatech.edu).

4.2. Results

4.2.1. Pre-experimental baseline pupil size

First, at the bivariate level we found that WMC explained 6% of the variance in baseline pupil size and with each 1 SD increase in WMC there was a 0.30 mm increase in baseline pupil diameter, b = 0.30, r = 0.24, $t(335)$ = 4.44, $p < 0.05$. Fluid intelligence explained 12% of the variance in baseline pupil size and each 1 SD increase in Gf was associated with a 0.45 mm increase in baseline pupil diameter, b = 0.45, r = 0.35, $t(335)$ = 6.90, $p < 0.05$. A hierarchical regression analysis, represented in Fig. 4, was conducted to test whether WMC predicted baseline pupil size above and beyond Gf. After controlling for Gf, there was no longer a significant relationship between WMC and baseline pupil size, $\Delta R^2 < 0.01$, $\Delta F(1, 334)$ = 0.01,

---

[2] In Studies 1 and 3 we analyzed pupil size as a function of memory load. However, task-related changes in pupil size in Study 3 were calculated and analyzed differently than they were in Study 1. The reason for this difference has to do with the difference in research questions. Our rationale for how they were analyzed in Study 1 are discussed in Footnote 1. In Study 3, the analysis on task-related changes in pupil size are independent of the main research question of the entire article. That is, is baseline pupil size related to important cognitive abilities? The analysis in Study 3 has more to do with how pupil dilation (as an indicator of mental effort) changes over increasing memory load and if this is related to working memory capacity or fluid intelligence. This analysis, unlike Study 1, requires correcting for initial differences in baseline pupil size. Additionally, we analyzed pre-trial baseline pupil size in Study 3, but not Study 1.

$p > 0.05$. Fluid intelligence, however, still predicted baseline pupil size after controlling for WMC, b = 0.45, $r_{partial} = 0.27$, $p < 0.05$. These results provide strong evidence that it is Gf, not WMC, which is uniquely related to baseline pupil size.[3]

As a supplementary analysis, we analyzed pre-experimental baseline pupil size as an extreme groups design. This allowed us to determine if we have the same effect size as in Studies 1 and 2, giving us some idea whether the effects of WMC in Study 3 are not just weaker, simply due to sampling error. Study 1 and 2 used a quartile and tertiary split on WMC respectively; in this analysis we will also use a quartile and tertiary split on WMC and Gf.

Using a quartile split, we found the effect size of WMC to be smaller than in Study 1. The difference between high and low WMC on baseline pupil size was about 0.65 mm, $t(164) = 4.06$, $p < 0.05$, $d = 0.63$. In study 1, there was a difference of about 0.97 mm ($d = 1.10$). The difference between high and low Gf on baseline pupil size was about 1.04 mm ($d = 1.07$).

Using a tertiary split, the difference between high and low WMC on baseline pupil size was about 0.58 mm, $t(222) = 4.15$, $p < 0.05$, $d = 0.56$. This is, in fact, almost identical to the effect size found in Study 2 (0.58 mm, $d = 0.53$). The difference between high and low Gf on baseline pupil size was about 0.83 mm, $t(222) = 6.24$, $p < 0.05$, $d = 0.84$. To compare the effect of Gf to that found in Study 2, we correlated Gf and baseline pupil size only using high and low WMC subjects. Gf correlated with baseline pupil size, $r(222) = 0.39$, $p < 0.05$. This is equivalent to that found in Study 2 (r = 0.37). These comparisons demonstrate the reliability of the effect sizes reported in all three studies.

Unsworth and Robison (2015) found that the effect of working memory load on phasic pupil dilations interacted with cognitive ability. However, they did not find any relationship between pre-trial baseline pupil size and their measure of working memory capacity (see K estimates from change detection tasks; Cowan et al., 2005). It is important to keep in mind that working memory capacity (K) as defined in Unsworth and Robison (2015) is not the same as WMC defined in the present study (Shipstead et al., 2014). Nevertheless, in this sample of subjects, we did also obtained estimates of K; although from a different and more challenging version than Unsworth and Robison (2015) (for a comparison see; Shipstead et al., 2014). Nevertheless, the two versions do have about the same predictive validity (Shipstead et al., 2014). In the current study, we found that K correlated with pre-experimental baseline pupil size (r = 0.22, $p < 0.05$) and pre-trial baseline pupil size (r = 0.244, $p < 0.05$). Individual differences in K predicted pre-experimental pupil size over and above WMC ($\Delta R^2 = 0.02$, $\Delta F(1,319) = 0.6.25$, $p < 0.05$) but not Gf ($\Delta R^2 < 0.01$, $\Delta F(1,334) = 0.47$, $p > 0.05$). Similarly, K predicted pre-trial baseline pupil size over and above WMC ($\Delta R^2 = 0.03$, $\Delta F(1,320) = 8.59$, $p < 0.05$) but not Gf ($\Delta R^2 < 0.01$, $\Delta F(1,320) = 2.00$, $p > 0.05$). However, Unsworth and Robison (2015) did find a relationship between variability of baseline pupil size and capacity (K). For variability in pre-experimental baseline pupil size we calculated the SD in pupil size over the 30-s measure. Variability in baseline pupil size did not correlate with WMC (r = 0.05, $p > 0.05$) and had a small correlation with Gf (r = 0.11, $p < 0.05$). Therefore, in our study we did find that baseline pupil size was related to capacity (K) and that variability in baseline pupil size had a small positive relationship to Gf. The discrepancy in our findings and those of Unsworth and Robison (2015) may be due to a number of factors, including; differences in task versions, sample characteristics, or the use of only one task to measure working memory capacity (reliability).

4.2.2. Controlling for potential confounds

Given that our procedure for subject recruitment is similar across the studies in our lab, there is the potential that some other confounding variables could explain our findings. Therefore, a number of these potential confounds will now be addressed.

There was a main effect of Ethnicity (Caucasians, African-Americans, and Other) on baseline pupil size, $F(2,328) = 9.57$, $p < 0.05$, $\eta^2_p = 0.06$. To rule out the effects of Ethnicity we analyzed the relationships between Gf and baseline pupil size in Caucasians (n = 62), African-Americans (n = 207), and Other (defined as *not* Caucasian or African-American, n = 62). Fig. 5 shows that Gf correlated with baseline pupil size in Caucasians, $r(60) = 0.34$, $p < 0.05$, African-Americans, $r(205) = 0.25$, $p < 0.05$, and Other, $r(60) = 0.38$, $p < 0.05$. A moderation analysis found that the relationship between Gf and baseline pupil size was not different across the Ethnicity groups, b = −0.06, t = −0.52, $p > 0.05$.

Using hierarchical regression analyses, we tested whether Gf can predict baseline pupil size even after controlling for a number of other demographic variables. Table 2 shows the relationship of 9 demographic variables to baseline pupil size. Only variables that had a significant relationship to baseline pupil size were included in the hierarchical regression; Ethnicity, Age, College Student, Nicotine, and Medication.

Together the demographic variables explained 15% of the variance in baseline pupil size, R = 0.39, $R^2 = 0.15$, $F(4,272) = 11.86$, $p < 0.05$. However, only Nicotine, b = −0.47, $r_{partial} = −0.15$, $p < 0.05$, and Medication, b = 0.56, $r_{partial} = 0.14$, $p < 0.05$, were unique predictors of baseline pupil size, which together accounted for 6% of the variance in pupil size. Even after controlling for the demographic variables, Gf still explained additional variance in baseline pupil size, $\Delta R^2 = 0.05$, $\Delta F(1,271) = 16.50$, $p < 0.05$. The unique relationship between baseline pupil size and Gf was relatively unaffected by the demographic variables, b = 0.32, $r_{partial} = 0.24$, t = 4.06, $p < 0.05$. Therefore, we have provided strong evidence that the relationship between Gf and baseline pupil size cannot be explained by these potentially confounding variables. There may be others we have not considered but these variables did not account for the Gf $\times$ baseline pupil size effect.

---

[3] We ran the same model, as Fig. 4, using a structural equation modeling approach. This approach uses latent factors rather than composite scores. The conclusion of the analysis is the same. The model overall was a good fit $\chi^2 (12) = 26.11$, $p < 0.05$, CFI = 0.98, RMSEA = 0.06. WMC did not predict pupil size (r = −0.06, $p > 0.05$). Gf did predict pupil size (r = 0.44, $p < 0.05$).
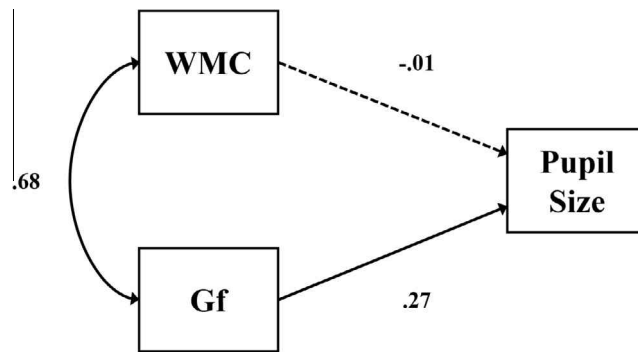
Fig. 4. Unique contributions of working memory capacity (WMC) and fluid intelligence (Gf) to baseline pupil size. $N = 337$. Path values are partial correlations.

4.2.3. Task-related pupil size

4.2.3.1. Pre-trial baseline pupil size. There was a main effect of WMC, $F(1,115) = 11.07$, $p < 0.05$, and Gf, $F(1,110) = 27.93$, $p < 0.05$, on pre-trial baseline pupil size (Fig. 6). However, there was no interaction between memory load and either WMC, $F(7,805) = 0.78$, $p > 0.05$, or Gf, $F(7,770) = 0.66$, $p > 0.05$. That is, the effect of WMC and Gf on baseline pupil size did not change as a function of performing the memory task. Additionally, there were small differences amongst the 8 trials on pre-trial baseline pupil size, $F(7,1708) = 4.82$, $p < 0.05$. This effect was not due to pre-trial baseline pupil size increasing with time on the task but rather a function of previous set-size. That is, pre-trial baseline pupil size was larger when the previous set-size was high. A similar finding was found in Heitz et al. (2008) and was attributed to the persistence of the phasic response from the previous trial. However, this may have resulted from a preparatory × previous set size interaction effect given the long ITI of 9 s.

4.2.3.2. Task-related pupil dilation. Consistent with findings in the literature, we found that pupil size increased as memory load increased, $F(11,3454) = 135.53$, $p < 0.05$. Whether pupil dilation plateaued or declined once memory capacity was reached depended on the cognitive ability level of the subject. As seen in Fig. 7a, both high WMC and high Gf subjects plateau whereas low WMC and low Gf subject's pupil dilations begin to decline around a memory load of 8 items (Trial × Ability interactions, $p < 0.05$). One potential explanation for this is that, high WMC/Gf individuals persist in giving their full mental effort beyond their capacity limits whereas low WMC/Gf individuals no longer give their full effort once their capacity limits have been reached. In fact, this has been observed in other studies from our lab (Engle, Cantor, & Carullo, 1992; Engle & Marshall, 1983).

4.3. Discussion

In Study 3, we tested whether working memory capacity and fluid intelligence have a unique relationship to baseline pupil size. We found that both working memory capacity and fluid intelligence correlated with baseline pupil size, however, only fluid intelligence uniquely predicted baseline pupil size (Fig. 4). Importantly, we also showed that a number of potential confounding variables such as ethnicity, age, or drug substances could not explain this relationship between fluid intelligence and baseline pupil size. This finding demonstrates limitations of using extreme group designs on cognitive abilities that are highly correlated with one another. Because working memory capacity and fluid intelligence are so highly correlated they are confounded when using an extreme groups design leading to limitations in interpreting the nature of the psychological and neurobiological mechanisms which distinguish them.

Similar to findings in Study 1 we also showed that baseline pupil size differences remained between high and low cognitive ability subjects while performing a memory-span task. If differences in mental effort could account for this finding, it would be expected that the baseline pupil size difference between high and low ability subjects would go away or perhaps decrease. However, this was not the case. There were small, if any differences, in the increase of pupil size as memory load increased until about 6–8 items (Fig. 7a). Interesting to note, there was a pupil dilation for the first memory item, which then plateaued until about four memory items at which point pupil dilation continues to increase. One potential explanation is that subjects were able to chunk the first 3 memory items. Not until about 8 memory items does the pattern of pupil dilations change for high and low ability subjects. While high ability subject's pupil size plateaued at about 8 memory items, low ability subject's pupil size declined. This suggests that low ability subjects no longer persist in mental effort once memory capacity is reached, indicated by accuracy rates at set-size of 9 and 12 (Fig. 7b).

5. General discussion

Throughout the three studies, we have shown that large differences in baseline pupil size, even observable to the unaided eye, exist between high and low cognitive ability individuals. Further, this difference does not go away when subjects are
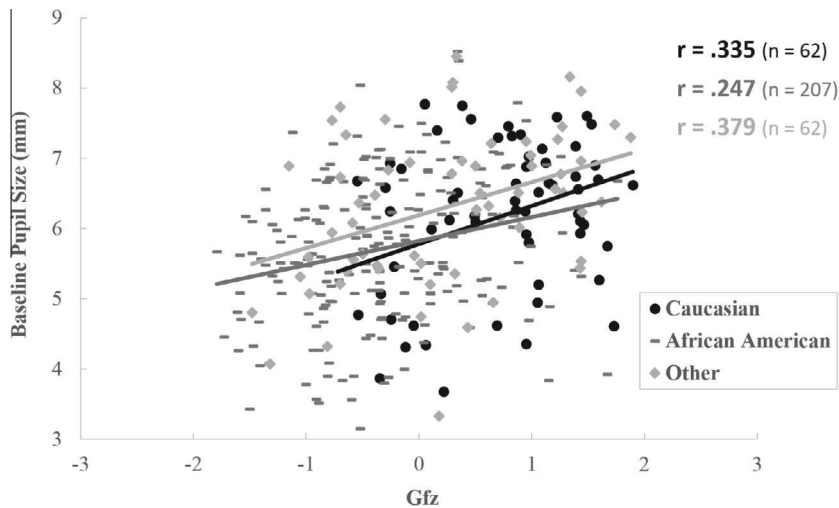
Fig. 5. Correlation between fluid intelligence (Gf) and baseline pupil size for Caucasians, African-Americans, and Other.

Table 2
Correlations between demographic variables and baseline pupil size.

Demographic	Relationship to baseline pupil size r(b)
Age (Years)	−0.32 (−0.07)[*]
College Student[a]	0.28 (0.60)[*]
Nicotine[b]	−0.21 (−0.69)[*]
Medications[c]	0.12 (0.56)[*]
Gender[d]	0.05 (0.12)
Handedness[e]	−0.05 (−0.14)
Caffeine[f]	−0.06 (−0.14)
Alcohol[g]	−0.03 (−0.12)
Sleep[h]	−0.07 (−0.05)
Gf (controlled)	**0.24 (0.34)**[*]

Note: For the demographic variables correlations are given outside parentheses and unstandardized coefficients are given inside parentheses.
   [a] Non-college vs. College student.
   [b] No-nicotine vs. Nicotine (in the last 10 h).
   [c] No-medications vs. Medications (that affect attention or memory in the last 24 h).
   [d] Male vs. Female.
   [e] Right vs. Left.
   [f] No-caffeine vs. Caffeine (in the last 8 h).
   [g] No-alcohol vs. Alcohol (2 or more drinks in the last 24 h).
   [h] Hours of sleep.
   [*] $p < 0.05$.

engaged in a cognitively demanding task and cannot be explained by differences in mental effort (Study 1). We found that the effect of working memory capacity and fluid intelligence on pupil size persisted even after multiple measurements and taking into account the effect of novelty or familiarity with the environment (Study 2). Finally, we found evidence that it is fluid intelligence, not working memory capacity, which is uniquely related to baseline pupil size (Study 3). Although some demographic variables, such as age and drug substances, were related to baseline pupil size they did not account for the relationship between Gf and baseline pupil size. Therefore, together these findings provide convincing evidence that the size of the pupil is a reliable and valid indicator of one's cognitive ability and intelligence. We believe that this finding is important for what it may reveal about the underlying neural mechanisms of fluid intelligence.

Although we did not obtain any direct neural measures of brain function, neuroscience research has shown a close association of pupil size with activity in the locus coeruleus-norepinephrine system (LC-NE) (Joshi et al., 2016; Murphy et al., 2014; Rajkowski et al., 1993; Varazzani et al., 2015). In fact, changes in pupil size have commonly been used as an indicator of locus coeruleus activity (Aston-Jones & Cohen, 2005b; Eldar, Cohen, & Niv, 2013; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Rajkowski et al., 1993). Locus coeruleus function can be characterized by two modes of activity. In a phasic mode, the locus coeruleus shows bursts of activity facilitating task-specific processes and task engagement; similar to the pupil dilation response studied by psychologists (Aston-Jones & Cohen, 2005b). A tonic mode of baseline locus coeruleus activity adjusts the gain of cortical circuits to disengage task-specific processes (Aston-Jones & Cohen, 2005b). Activity in the locus coeruleus
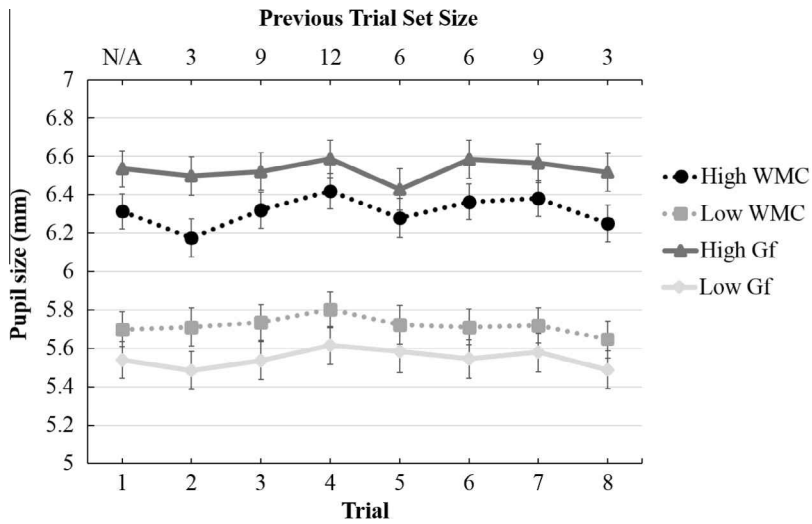
Fig. 6. Pre-trial baseline pupil size on each trial for high WMC/Gf and low WMC/Gf subjects.
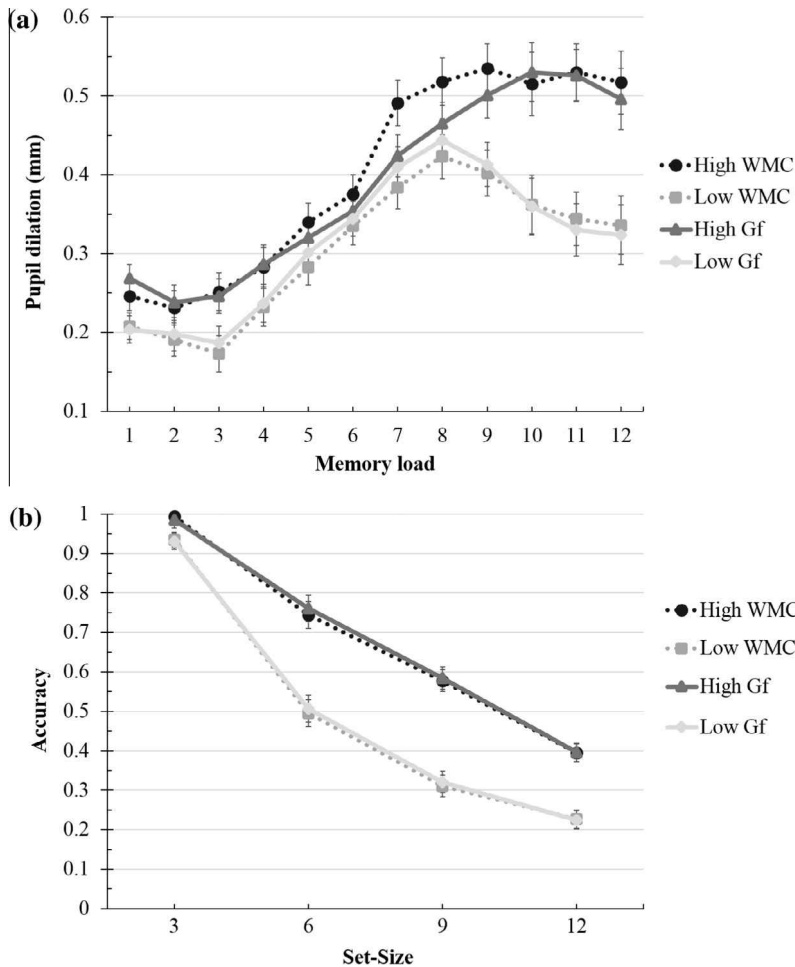

Fig. 7. (a) Pupil dilation as a function of memory load for high WMC/Gf and low WMC/Gf subjects. (b) Proportion accurate as a function of set-size for high WMC/Gf and low WMC/Gf subjects.

releases norepinephrine and modulates the gain of target neurons to be more sensitive to incoming signals (both excitatory and inhibitory) (Berridge & Waterhouse, 2003). Because of the widespread projections the LC-NE system has, this modulation of neural gain has an effect on the strength of functional connectivity throughout the brain (Moore & Bloom, 1979). In fact, researchers have even shown that changes in pupil size (as an index of tonic locus coeruleus activity) corresponds to this effect of the LC-NE system on global levels of functional connectivity in the human brain (Eldar et al., 2013; Warren et al., 2016).

Nonetheless, in and of itself, this does not completely explain why in the absence of performing a cognitive task, during a passive baseline, we would find differences in pupil size related to intelligence. To better understand our findings we need to consider what goes on in the resting-state brain during a passive baseline condition. Seminal findings by Raichle and colleagues (Raichle et al., 2001; Shulman et al., 1997) lead to the discovery that the resting-state brain exhibits an intrinsic functional organization even during a passive baseline condition (Raichle, 2015a). It was first discovered that a network of brain regions were consistently showing *deactivation* during performance on a wide variety of cognitive tasks, relative to a passive baseline condition (Shulman et al., 1997). It is now well established that this network of brain regions, dubbed the default-mode network, is not 'active' in the baseline condition but rather the fluctuations in baseline levels of activity form a functional organization (Fox et al., 2005; Raichle et al., 2001).

Studies on the resting-state brain suggest there exists a functional balance between default-mode network regions and other brain networks, and that this balance has implications for cognition and behavior (Raichle, 2015b). For instance, there exists an antagonistic relationship between default-mode and executive attention networks (Raichle, 2015b). A greater deactivation in default-mode regions, and thereby less interference from this network, as one engages in an attention demanding task is related to better performance (Sambataro et al., 2010). Also, in the resting-state brain, the default-mode network and executive network regions are negatively correlated with one another (Fox et al., 2005).

Functional organization in the default-mode network is a strong candidate to explain our findings due to its central importance to organization in the resting-state brain (Raichle, 2015a). However, there are also empirical reasons we believe our findings on baseline pupil size is related to the default-mode network. As already discussed above, the locus coeruleus-norepinephrine system has an effect on global levels of functional connectivity, and pupil size corresponds to this effect on functional connectivity (Eldar et al., 2013; Warren et al., 2016). It has also been shown that fluctuations in baseline pupil size correlates with activity in default-mode regions during a passive baseline condition (Yellin, Berkovich-Ohana, & Malach, 2015). Researchers are also beginning to show that cognitive ability correlates with resting-state functional connectivity. The anti-correlation between default-mode and executive networks is related to individual differences in working memory capacity (Keller et al., 2015), which is highly correlated with Gf. Also, recent findings from the Human Connectome Project also show that functional connectivity patterns in default-mode and frontoparietal areas are associated with fluid intelligence (Finn et al., 2015; Smith et al., 2015); though only measured using a single task, the Ravens Progressive Matrices.

All these findings, together with a theoretical understanding of the role the LC-NE system plays in cognitive functioning, suggests a relationship amongst baseline pupil size, tonic locus coeruleus activity, and resting-state functional connectivity that gives rise to individual differences in fluid intelligence. *Therefore, we believe 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.* Specifically, that fluid intelligence is related to larger baseline pupil size which indicates stronger functional connectivity in default-mode and executive attention networks, arising from the neuromodulatory role of the locus coeruleus system.

This stronger functional connectivity may allow higher fluid intelligence individuals to disengage from a default-mode more quickly and to activate and maintain a task-focused, goal-oriented mode more easily. This is consistent with the adaptive gain theory of locus coeruleus function whereby the locus coeruleus is involved in the dynamic balance between states of task engagement and disengagement (Aston-Jones & Cohen, 2005b). Consistent with the Yerkes-Dodson curve (Yerkes & Dodson, 1908), low levels of tonic locus coeruleus activity is associated with being non-alert and poor performance, moderate levels of tonic activity is associated with optimal performance and phasic activity, whereas too high levels of tonic activity is associated with distractibility and poor performance (Aston-Jones & Cohen, 2005b).

Activation in default-mode network has been implicated in supporting processes for mind wandering, an instance of task disengagement (Christoff, 2012; Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Mason et al., 2007). The occurrence of mind wandering while performing a cognitive task is thought to be due to failures to maintain task-focused, goal-oriented attention and is negatively correlated with working memory capacity, attention control, and fluid intelligence (McVay & Kane, 2009, 2010; Robison, Gath, & Unsworth, 2016; Smallwood, 2013; Unsworth & McMillan, 2014). Furthermore, a recent study has found that different states of off-task attention lapses correspond to changes in baseline pupil size as would be predicted by tonic locus coeruleus activity (Unsworth & Robison, 2016). States of low arousal, such as inattentiveness and mind wandering, were associated with smaller baseline pupil size (low tonic activity); states of on-task focus was associated with moderate baseline pupil size (moderate tonic or phasic activity); and states of distractibility, defined as external distraction, was associated with larger baseline pupil size (high tonic activity). The dynamic relationship between default-mode and executive control networks may be associated with the adaptive adjustments of tonic and phasic modes of locus coeruleus activity. Future research is needed to better understand how, and if, the dynamic relationships amongst functional connectivity of brain networks, locus coeruleus activity, and task disengagement is related to higher order cognitive abilities.

What we have shown is that individual differences in fluid intelligence is related to differences in baseline pupil size. However, at this point, our brain story of the intelligence - baseline pupil size relationship is only reasonably informed

speculation. Further research is needed to follow up on our findings if we want to draw any definite conclusions about the underlying neural mechanisms. By understanding the different mechanisms of working memory capacity and fluid intelligence we can better understand what gives rise to individual differences in these abilities, how they are different from one another, and potentially how they can be improved. Therefore, further investigation of the neurobiological processes reflected in individual differences in baseline pupil size is warranted.

Acknowledgments

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30–60. http://dx.doi.org/10.1037/0033-2909.131.1.30.

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

Aston-Jones, G., & Cohen, J. D. (2005a). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *Journal of Comparative Neurology, 493*(1), 99–110. http://dx.doi.org/10.1002/cne.20723.

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

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

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge University Press.

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. http://dx.doi.org/10.1016/S0165-0173(03)00143-7.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Christoff, K. (2012). Undirected thought: Neural determinants and correlates. *Brain Research, 1428*, 51–59. http://dx.doi.org/10.1016/j.brainres.2011.09.060.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106*(21), 8719–8724. http://dx.doi.org/10.1073/pnas.0900234106.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (1983). Applied multiple regression/correlation for the behavioral sciences.

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. http://dx.doi.org/10.1016/j.tics.2003.10.005.

Cowan, N., Elliott, E. M., Saults, S. J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42–100. http://dx.doi.org/10.1016/j.cogpsych.2004.12.001.

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton NJ Educational Testing Service (Vol. 102). doi:http://dx.doi.org/10.1073/pnas.0506897102.

Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*. http://dx.doi.org/10.1038/nn.3428.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972–992.

Engle, R. W., & Marshall, K. (1983). Do developmental changes in digit span result from acquisition strategies? *Journal of Experimental Child Psychology, 36*, 429–436.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309–331. http://dx.doi.org/10.1037/0096-3445.128.3.309.

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*. http://dx.doi.org/10.1038/nn.4135.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673–9678. http://dx.doi.org/10.1073/pnas.0504136102.

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. http://dx.doi.org/10.3758/CABN.10.2.252.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*. http://dx.doi.org/10.1111/j.1469-8986.1996.tb01071.x.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2014). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition, 1–8*. http://dx.doi.org/10.3758/s13421-014-0473-3.

Heitz, R. P., Redick, T. S., Hambrick, D. Z., Kane, M. J., Conway, A. R. A., & Engle, R. W. (2006). Working memory, executive function, and general fluid intelligence are not the same. *Behavioral and Brain Sciences, 29*(02), 135–136.

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. http://dx.doi.org/10.1111/j.1469-8986.2007.00605.x.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*(3423), 349–350.

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

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

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*(January), 221–234. http://dx.doi.org/10.1016/j.neuron.2015.11.028.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences, 30*, 135–187. http://dx.doi.org/10.1017/S0140525X07001185.

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

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637–671. http://dx.doi.org/10.3758/BF03196323.

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

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. http://dx.doi.org/10.1016/j.cortex.2014.12.001.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wengner, D. M., Grafton, S. T., & Macrae, N. C. (2007). Wandering minds: The default network and stimulus-independent thought. *Science, 315*(5810), 393–395. http://dx.doi.org/10.1126/science.1131295.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 196–204. http://dx.doi.org/10.1037/a0014104.

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin, 136*(2), 188–197. http://dx.doi.org/10.1037/a0018298.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.

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. http://dx.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. http://dx.doi.org/10.1002/hbm.22466.

Peavler, S. W. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*(5), 559–567.

Raichle, M. E. (2015a). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433–447. http://dx.doi.org/10.1146/annurev-neuro-071013-014030.

Raichle, M. E. (2015b). The restless brain: How intrinsic activity organizes brain function. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 370*(1668), 20140172. http://dx.doi.org/10.1098/rstb.2014.0172.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676–682. http://dx.doi.org/10.1073/pnas.98.2.676.

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* (Vol. 19, pp. 974). .

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

Robison, M. K., Gath, K. I., & Unsworth, N. (2016). The neurotic wandering mind: An individual differences investigation of neuroticism, mind-wandering, and executive control. *The Quarterly Journal of Experimental Psychology*. http://dx.doi.org/10.1080/17470218.2016.1145706.

Sambataro, F., Murty, V. P., Callicott, J. H., Tan, H.-Y., Das, S., Weinberger, D. R., & Mattay, V. S. (2010). Age-related alteration in default mode network: Impact on working memory performance. *Neurobiology of Aging, 31*(5), 839–852. http://dx.doi.org/10.1016/j.neurobiolaging.2008.05.022.Age-related.

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. http://dx.doi.org/10.1016/j.jml.2014.01.004.

Shulman, G. L., Fiez, J. a., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience, 9*(5), 648–663. http://dx.doi.org/10.1162/jocn.1997.9.5.648.

Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity. *Psychological Bulletin, 139*(3), 519–535. http://dx.doi.org/10.1037/a0030010.

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., J 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*. http://dx.doi.org/10.1038/nn.4125.

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

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. http://dx.doi.org/10.3758/BF03192720.

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

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. http://dx.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*. http://dx.doi.org/10.3758/s13415-016-0417-4.

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. http://dx.doi.org/10.1111/j.1469-8986.2009.00884.x.

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *The Journal of Neuroscience, 35*(20), 7866–7877. http://dx.doi.org/10.3389/fnbeh.2015.00310.

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.

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. http://dx.doi.org/10.1016/j.neuroimage.2014.11.034.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.