

Is baseline pupil size a good measure to assess attentional control?

Dahae Cho (dc5836@nyu.edu)

Department of Biotechnology and Entrepreneurship, New York University, USA

Sunwoo Moon (moons@hanyang.ac.kr)

Department of Cognitive Sciences, Hanyang University, Seoul, Republic of Korea

Hyungwook Yim (hwyim@hanyang.ac.kr)

Department of Cognitive Sciences, Hanyang University, Seoul, Republic of Korea

Abstract

Research has indicated that baseline pupil size may reflect attentional control capabilities. Several studies have demonstrated links between resting pupil measurements and various cognitive functions, including working memory capacity and fluid intelligence. The existing literature presents mixed evidence, with some studies supporting this relationship while others fail to replicate these findings. Moreover, the validity of this relationship across different populations has come into question, particularly when considering age-related physiological changes that affect pupil size. Our study sought to investigate whether the relationship between baseline pupil size and attentional control remains consistent across age groups. Results showed no meaningful correlation between baseline pupil size and attentional control. In examining possible explanations for these disparate results, we identified several methodological challenges, including inconsistencies in testing environments and variations in pupil measurement protocols, that may account for the conflicting findings in previous research. These observations highlight the importance of developing more standardized experimental approaches to properly evaluate baseline pupil size as an indicator of attentional control, particularly when studying aging populations.

Keywords: baseline pupil size; working memory capacity; LC-NE system; attentional control

Introduction

Attention has two main aspects: mental effort and selective processing. Attention as mental effort assumes that the central processing mechanism has limited capacity. This mechanism is associated with conscious and controlled processing and determines the limits of divided attention - the extent to which different information sources can be processed simultaneously. However, there are various disagreements about this perspective, including rejecting the concept of limited capacity itself or assuming multiple pools of limited resources. On the other hand, there is no disagreement about the selective nature of processing (Johnston, W.A., & Dark, V.J., 1986).

Environmental information processing is divided into bottom-up (or data-driven) and top-down (or internally-driven) processing. In bottom-up processing, stimuli activate various levels of encoding, from simple physical and sensory analysis to complex semantic and schematic analysis. The depth of processing is determined by the

intensity and clarity of the stimulus, as well as the availability of codes formed through learning. When stimuli are weak, noisy, or unfamiliar, stimulus processing becomes data-limited and remains at shallow levels of analysis.

This selective attention mechanism plays a crucial role in improving information processing efficiency and effectively allocating cognitive resources. However, as age increases, the efficiency of these attention mechanisms decreases, leading to various memory problems. In particular, it becomes more difficult to remember event details, and source memory errors occur frequently (Chen & Naveh-Benjamin, 2012; Boywitt, Kuhlmann, & Mesiser, 2012). The causes of this memory decline include cognitive decline in the brain, decreased ability to bind event components, reduced inhibition of distracting stimuli, and difficulties in suppressing interference from previous experiences.

Studies have shown how these phenomena occur and how attention can be both focused and divided simultaneously. Additionally, selective attention function is closely related to frontal lobe maturation (Squire et al., 2013; Diamond, 2002), and the relatively late development of the frontal lobe provides important insights for developmental research. For example, studies comparing category learning between infants and adults show that while both groups can learn categories, only adults selectively attend to category-relevant dimensions and optimize attention during the learning process (Best, Yim, & Sloutsky, 2013). In contrast, infants tend to distribute attention across multiple dimensions.

The biological basis of these attentional control processes is closely linked to the LC-NE (Locus Coeruleus-Norepinephrine) system. Located in the brainstem, this system releases norepinephrine throughout the cerebral cortex, strengthening goal-related behavior while inhibiting irrelevant behavior (Berridge & Waterhouse, 2003; Aston-Jones & Cohen, 2005). According to the Arousal-Biased Competition theory, this system mediates global arousal in the brain and influences cognitive processes through its widespread efferent projections. When arousal is induced, it can either elevate or impair an object's neural representation based on its given priority, creating a "winner-take-more" and "loser-take-less" selection effect in perception and memory (Mather & Sutherland, 2011). The Glutamate

Amplifies Noradrenergic Effect (GANE) model suggests that this prioritization occurs through local noradrenergic “hot spots” where the co-release of glutamate and norepinephrine signals elevated attentional priority. Importantly, noradrenaline release outside these hot spots produces an inhibitory effect, explaining arousal-induced suppression phenomena.

The LC-NE system operates in two modes: tonic activity regulating basic arousal levels and phasic activity responding to important events, which has been proposed as an important mechanism explaining individual differences in attentional control (Unsworth & Robison, 2017a). This system’s functionality significantly changes with age. Older adults show reduced locus coeruleus MRI contrast compared to young adults, indicating changes in neuronal density. This reduction is associated with decreased cognitive function and increased risk for neurodegenerative diseases. Research suggests that older adults experience noradrenergic hyperactivity or elevated tonic noradrenergic firing rates, which leads to decreased phasic noradrenaline release and sub-optimal performance. In contrast, young adults with unaltered tonic noradrenergic release show improved task performance, speed, and perceptual contrast under increased arousal conditions.

Recent studies have focused on pupil diameter as a non-invasive measure of LC-NE system activity. Correlations between resting state pupil size and various attentional control tasks, including working memory capacity and fluid intelligence, have been reported (Tsukahara et al., 2016), suggesting that pupil size could function as a biological marker of cognitive ability.

In this context, the effects of aging on both pupillary responses and attentional control deserve attention. Aging involves various biological changes including decreased pupil diameter, reduced flexibility, and altered responses to light and cognitive load, occurring simultaneously with cognitive decline in working memory and attention control domains. This raises important questions about the validity of pupillometry as a cognitive measurement tool in elderly populations. This study examines whether pupil diameter remains a reliable indicator of attentional control across the lifespan, considering both the multifaceted nature of attentional control and age-related changes. In particular, it focuses on the validity of pupillometry in elderly populations and investigates how age-related changes in pupillary response patterns interact with various components of selective attention.

Experiment

Methods

Participants

118 participants (59 younger adults: 31 females, 28 males, $M = 21.02$ yrs, $SD = 1.82$ yrs/ 59 older adults: 29 females, 30 males, $M = 64.18$ yrs, $SD = 2.49$ yr) younger adults were recruited through an online community page of Hanyang

University, Seoul, Republic of Korea, and older adults were recruited through Gallup Korea. The experiments described in this paper were conducted as part of a larger research project investigating human memory aging. Specifically, these experiments represent the first day of data collection in this comprehensive project, and the study was conducted over 7 weeks and consisted of two phases (Phase 1 and Phase 2). This paper focuses on a subset of the complete experimental protocol conducted during the project’s initial session. Also, the college student participants were compensated with 100,000 KRW (approximately 68 USD) for their participation, and the elderly participants were compensated with 200,000 KRW (approximately 136 USD). The research was approved by the Institutional Review Board at Hanyang University.

Materials, Design & Procedure

The order of experiments was fixed as follows: (1) pupil measurement, (2) simple episodic memory task (list length manipulation), (3) simple reaction task, (4) anti-saccade task, (5) choice reaction task, (6) symmetry span task, (7) color-shape switch task, (8) complex memory binding task. The two memory tasks (2, 8) were not measured for another study, which will be reported elsewhere.

Pupil Measurement: Participants were asked to fix their forehead and chin on a chin/head rest and adjust their chair height. The participant’s dominant eye was used for the measurement. After a 9-point calibration procedure, pupil size was measured in two stages. First, resting state pupil size was recorded for 100 seconds, with participants fixating on a central black cross. The lights in the room were turned off, and the background of the screen was mid-grey (RGB; 220, 220, 220). This generated a luminance of 12 lux from the participant’s view, as suggested by Tsukahara & Engle (2021). A 30-second fake eye recording followed to record a reference point in order to convert arbitrary values generated by the eye-tracker into actual size in millimeter units. Pupil size was measured using an Eye Link 1000 Plus eye tracker (SR Research Ltd., Canada). A tower-mount configuration with a chin rest and forehead support minimized head movement, with the sampling rate set at 1000Hz.

Simple Reaction Task: Participants encountered a letter ‘X’ within a square frame, appearing at variable intervals (1-3 seconds). Their task was to respond via the spacebar immediately upon detecting the ‘X’. The task structure included eight practice attempts followed by 20 experimental trials, based on Deary et al.’s (2010) design.

Antisaccade task: Participants viewed a white fixation cross against a gray background, displayed for either 1 or 2 seconds. Following this, a white asterisk flashed at a 12.3° visual angle on either the screen’s left or right side for 300ms. Subsequently, either the letter O or Q briefly appeared (100ms) on the opposite side before being masked by ‘##.’ With participants positioned 60cm from the display, their task was to direct their gaze opposite to the asterisk’s location and identify the letter within a 5-second window.

Performance feedback was displayed for 1 second using cyan ‘O’ for correct and magenta ‘X’ for incorrect responses. The task began with 16 practice trials, followed by two experimental blocks of 36 trials each, following Draheim, Tshukara, and Engle’s (2023) protocol.

Choice Reaction Task: The display featured four white squares arranged horizontally at the screen center, with each square corresponding to specific keyboard letters (x, c, b, n). Participants were instructed to press the appropriate key when a black ‘X’ appeared in any square, with responses required as quickly as possible. Intervals between trials varied from 1 to 3 seconds. Following eight practice trials, participants completed 40 experimental trials, adhering to Deary et al.’s (2010) protocol.

Symmetry span task: The task involved memorizing the sequential positions of red squares within a 4×4 grid, with an intervening distraction activity between square presentations. The distraction consisted of evaluating whether black and white patterns displayed on a 16×16 grid exhibited left-right symmetry. To ensure task comprehension, participants completed three distinct training sessions. During the initial training, participants viewed only the sequential presentation of red squares, completing four sequences (two sequences each of both two and three squares). The second training focused exclusively on symmetry assessment, comprising 15 trials with immediate post-response feedback (1-second duration). The response time threshold for subsequent phases was established as the mean response time plus 2.5 standard deviations from this second training session. The third training phase combined both elements: participants evaluated symmetry patterns within their individualized time constraint (the distraction component), followed by a 650msec presentation of a red square whose location needed to be remembered (the primary memory task). Participants were notified about the time restrictions for symmetry evaluations, and if they exceeded their time limit, the red square appeared immediately. This final practice included three sequences of two squares each, with participants required to maintain at least 85% accuracy on the symmetry evaluations. The main experimental trials followed the same protocol as the final practice but expanded the sequence lengths to include between two and five squares, with three repetitions at each sequence length. This experimental paradigm was adapted from Unsworth and Brewer’s (2009) symmetry span methodology.

Color-Shape Switch task: Each trial began with a brief (150ms) presentation of either \triangle or \square (Korean consonants representing ‘color’ and ‘shape’ respectively), followed by both the cue and a colored geometric figure below. The figure was either a triangle or a circle, displayed in either red or green. When presented with \triangle , participants identified the color (red/green), while \square prompted shape identification (triangle/circle) using keyboard inputs (Z/M). Trials were separated by 2.5s intervals, with 1-second feedback periods. Training consisted of two 12-trial blocks,

while the main task comprised two 24-trial blocks. The sequence included 24 non-switch and 23 switch trials, designed to avoid more than four consecutive switches or non-switches. This paradigm was based on Miyake et al.’s (2000) methodology.

Behavioral Results

For blink detection, we utilized the PuPL program, implementing the algorithm ‘Identify blinks using pupillometry noise’ (Hershman, Henik, & Cogen, 2018). We excluded data points 50ms before and 150ms after each detected blink to ensure data quality and remove invalid data due to blinks. This approach aligns with standard preprocessing procedures, as pupil size data during blinks is typically unreliable and requires careful handling. The blink detection and removal process is particularly crucial as pupil-size data, even when of high quality, invariably contains missing and invalid data points. Our preprocessing approach focused on first identifying invalid data and removing it, effectively converting invalid data into missing data.

The measurements, initially recorded in arbitrary units, were calibrated to millimeters. The mean baseline pupil size for young adults was ($M = 4.47\text{mm}$ ($SD = .70$)) was larger than the mean baseline pupil size for older adults ($M = 3.86\text{mm}$, $SD = .72$; independent t-test, $t(116) = 4.64$, $p < .001$; see Figure 1). The result is consistent with known age-related changes in pupil-related response patterns. This difference in baseline pupil size between age groups is particularly noteworthy as it may reflect underlying changes in the LC-NE system and could potentially impact the interpretation of pupillometric data in cognitive studies involving different age groups.

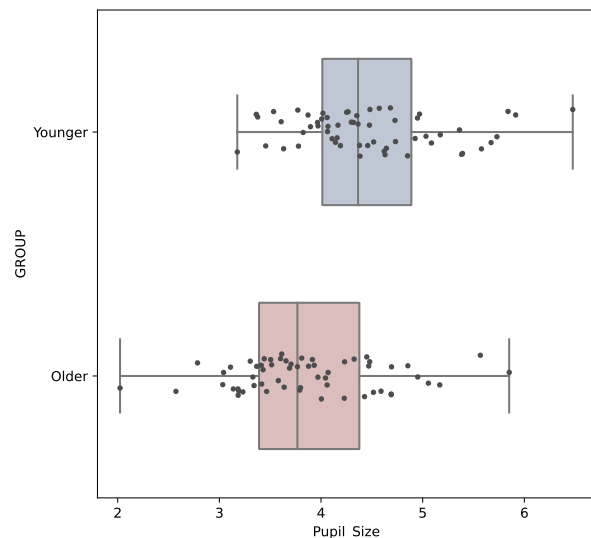


Figure 1: Distribution of baseline pupil size for younger and older adults. The scatter plot illustrates individual differences in pupil size across age groups, with younger adults (top) and older adults (bottom) showing distinct distributions.

Correlations between baseline pupil size and cognitive ability

In order to examine whether pupil size can serve as a proxy for attentional control, we calculated Pearson correlation coefficients between baseline pupil size and five attention control tasks for older and younger adults respectively. For the Anti-saccade task accuracy was used, for the Color-Shape task reaction time was used, for the Symmetry span task accuracy was used, and for the Simple reaction and Choice reaction task reaction time was used. Pearson correlation coefficients were statistically tested using both frequentists (i.e., p-value) and Bayesian approaches (i.e., Bayes factor) using JASP (JASP Team, 2023). We especially calculated Bayes factors (BF_{01}) to evaluate the evidence for the null-hypothesis. We used Jeffreys's (1961) interpretation for Bayes factor analysis. The interpretation of *and* BF_{01} values are based on thresholds of 1, 3, and 10: values between 1 and 3 indicate weak support for the null hypothesis, values between 3 and 10 indicate substantial support and values above 10 indicate strong support.

Results showed weak correlations across all tasks when examining Pearson's correlation coefficients, indicating no relationship between baseline pupil size and task performance (see Table 1). All coefficients were not statistically significant based on p-values. The Bayesian analysis showed that while older adults consistently demonstrated higher *and* BF_{01} values compared to younger adults across all tasks, both age groups showed evidence supporting the null hypothesis. This suggests that there is more support for the absence of a relationship between baseline pupil size and task performance in both groups, with this support being stronger in the older adult group. Results consistently indicate that there is no significant relationship between baseline pupil size and performance on

any of the five attention control tasks, with this pattern being particularly pronounced in the older adult group (see Table 1).

Discussion

The current study examined whether baseline pupil size can serve as a reliable indicator of attention control, considering the effects of aging. We implemented an experiment incorporating attention control tasks while measuring baseline pupil size across younger and older adults. Our results showed significant differences in baseline pupil size between younger and older adults. However, we found no significant correlations between baseline pupil size and attention control tasks. These findings challenge the proposed relationship between baseline pupil size and cognitive function, especially when accounting for age-related differences in attentional mechanisms. Results suggest that this physiological measure may not be a reliable indicator of cognitive ability as previously suggested, specifically when comparing across different age groups.

Our experiments did not find a significant correlation between baseline pupil size and working memory capacity, which aligns with several recent findings (Aminihajibashi, Hagen, Foldal, Laeng, & Espeseth, 2019; Unsworth et al., 2019; Unsworth & Robison, 2015, 2017b). The difficulty in finding a remarkable correlation between baseline pupil size and working memory capacity can be attributed to several key factors.

First, methodological variations across studies have likely contributed to inconsistent findings. Studies have used different tools to measure WMC (Working Memory Capacity), ranging from complex span tasks with concurrent information-processing demands to simpler tasks like Letter-Number Sequencing (LNS). While both assess

Tasks	Value	Older Adults	Younger Adults
Anti-saccade task (Accuracy)	Pearson's r BF_{01} (p -value)	-0.04 5.94 (0.79)	-0.09 4.94 (0.50)
Color-Shape switch task (RT)	Pearson's r BF_{01} (p -value)	0.02 6.09 (0.89)	-0.03 5.98 (0.81)
Symmetry span task (Accuracy)	Pearson's r BF_{01} (p -value)	-0.14 3.64 (0.30)	-0.20 2.06 (0.14)
Simple reaction task (SRT)	Pearson's r BF_{01} (p -value)	0.06 5.65 (0.67)	-0.13 3.82 (0.32)
Choice reaction task (CRT)	Pearson's r BF_{01} (p -value)	0.02 6.10 (0.90)	0.13 3.90 (0.33)

Table 1: Pearson correlation coefficients between baseline pupil size and attention control tasks across older and younger adults.

similar cognitive processes, the difference in task complexity may affect the sensitivity of the observed relationships. Draheim, Tshukara, & Engle (2024) have pointed out that a major factor that makes attention control tasks unreliable (or unstable) is due to its measurement relying on reaction time. However, the two tasks (i.e., anti-saccade, and symmetry span tasks) did not use reaction time as a measure, and the procedure was identical to the tasks that Engle and colleagues used.

Second, the duration of baseline pupil diameter measurement varies significantly across studies, from brief 7-30 second intervals to extended 5-minute periods in studies. Longer measurement durations may introduce greater variability due to time-on-task effects and fluctuations in participants' focus and arousal levels.

Third, environmental conditions, particularly lighting and screen luminance, have varied considerably across studies. Tsukahara et al. (2016) showed that settings with higher environmental luminance result in larger baseline pupil diameters, restricting the range and variability of the individual pupil size difference. However, the room setting (i.e., lights off) and luminance setting in our study (i.e., 12 lux) were congruent with the suggestion made by Tsukahara et al. (2016).

Fourth, cultural and motivational differences may influence results. Different participant populations (e.g., Norwegian versus American students) may exhibit varying levels of motivation and engagement during cognitive tasks, affecting pupil size and arousal levels. Additionally, individual differences in motivation, alertness, and interest during experimental sessions might influence baseline pupil size, though subsequent studies have found no significant relationship between these self-reported factors and baseline pupil diameter.

Meta-analytic findings suggest (Unsworth et al., 2021) that the overall correlation between WMC and baseline pupil size is near zero and they are not statistically significant ($r =$

.01, 95% CI [-.03, .06], $p = .63$). The consistent variability across studies suggests that other factors that are not considered in the analysis, such as lab-specific protocols or participant demographics, might be responsible for the mixed findings in our research.

Furthermore, cognitive states and mind-wandering during baseline measurement may confound results. Participants often report diverse thoughts during baseline measurement, ranging from mundane topics to emotionally charged concerns. These mental activities may influence arousal levels and pupil size, potentially undermining the assumption that baseline measures solely reflect tonic LC-NE system activity.

Finally, compared to previous studies that find a relationship between baseline pupil size and attentional control tasks (e.g., Draheim, et al., 2021), our study examines a smaller number of participants. In Draheim et al. (2021) there were more than 400 participants while we have only examined 50 participants. Therefore, it is possible that the current study lacked statistical power in capturing the relationship. However, if detecting the relationship between baseline pupil size and attentional control tasks requires a large sample size, it is questionable whether baseline pupil size can be a good proxy for measuring attentional control in practice.

These challenges underscore the complexity of studying the relationship between attentional control and baseline pupil size and highlight the need for future research with standardized protocols, larger and more diverse samples, and better-controlled conditions. Future studies should address these methodological and sampling issues. This will help clarify whether a robust relationship between baseline pupil size and attentional control truly exists.

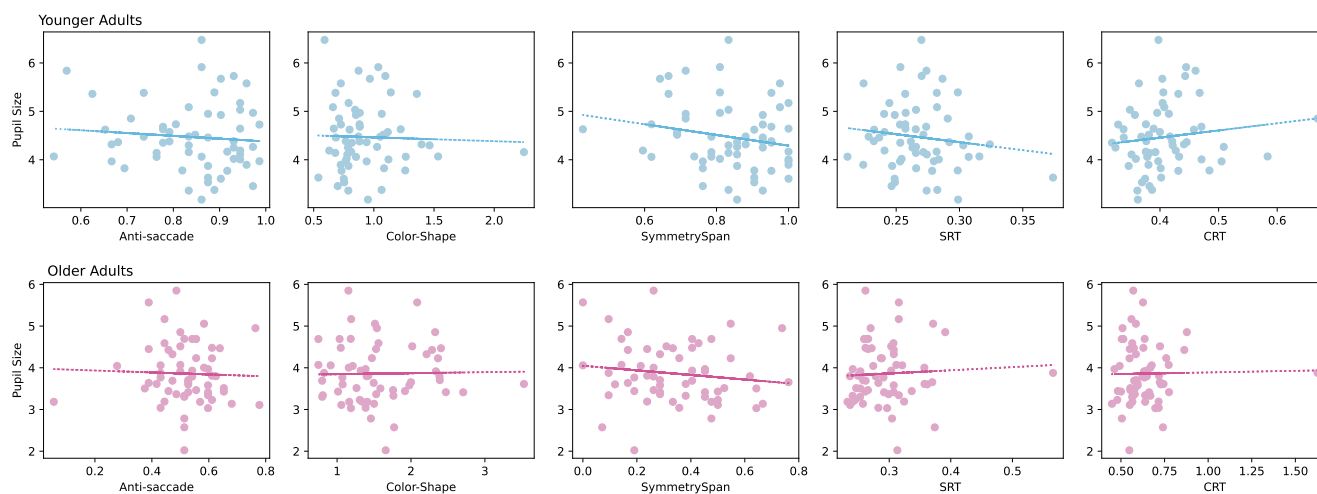


Figure 2: Relationship between baseline pupil size and attention control tasks for younger and older adults.

Acknowledgments

This work has been supported by the National Research Foundation of Korea funded by the Ministry of Science and ICT (No. 2018R1A5A7059549). Correspondence concerning this article should be addressed to Hyungwook Yim (E-mail: hwyim@hanyang.ac.kr)

References

- Aminihajbashi, 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.
- Aston-Jones G, Cohen JD. (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*:403-50.
- 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*(1), 33–84.
- Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology, 116*(2), 105–119.
- Boywitt, C. D., Kuhlmann, B. G., & Meiser, T. (2012). The role of source memory in older adults' recollective experience. *Psychology and Aging, 27*(2), 484–497.
- Chen, T., & Naveh-Benjamin, M. (2012). Assessing the associative deficit of older adults in long-term and short-term/working memory. *Psychology and Aging, 27*(3), 666–682.
- Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of experimental psychology. General, 150*(2), 242–275.
- Draheim, C., Tshukara, J. S., & Engle, R. W. (2024). Replication and extension of the toolbox approach to measuring attention control. *Behavior research methods, 56*(3), 2135–2157.
- Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37*, 1–26.
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron, 89*(1), 221–234.
- Lu, H., Van Der Linden, D., & Bakker, A. B. (2023). Changes in pupil dilation and P300 amplitude indicate the possible involvement of the locus coeruleus-norepinephrine (LC-NE) system in psychological flow. *Scientific Reports, 13*(1), 1908.
- Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods, 55*(6), 3055–3077.
- Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*(2), 114–133.
- Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics, 81*(2), 407–419.
- Robison, M. K., Coyne, J. T., Sibley, C., Brown, N. L., Neilson, B., & Foroughi, C. (2022). An examination of relations between baseline pupil measures and cognitive abilities. *Psychophysiology, 59*(12), e14124.
- Tsukahara, J. S., Draheim, C., & Engle, R. W. (2021). Baseline pupil size is related to fluid intelligence: A reply to. *Cognition, 215*, 104826.
- Tsukahara, J. S., & Engle, R. W. (2021). Fluid intelligence and the locus coeruleus-norepinephrine system. *Proceedings of the National Academy of Sciences, 118*(46), e2110630118.
- Tsukahara, J. S., & Engle, R. W. (2021). Is baseline pupil size related to cognitive ability? Yes (under proper lighting conditions). *Cognition, 211*, 104643.
- Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology, 91*, 109–123.
- Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective, & Behavioral Neuroscience, 19*, 100–113.
- Unsworth, N., Miller, A. L., & Robison, M. K. (2021). Is working memory capacity related to baseline pupil diameter? *Psychonomic Bulletin & Review, 28*(1), 228–237.
- Unsworth N, Robison MK.(2017a). A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control. *Psychon Bull Rev. 2017 Aug;24(4)*:1282-1311.
- 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.
- Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*(6), 2005–2015.
- Yim, H., & Yoon, S. (2024). Is focusing enough in category learning? In L. K. Samuelson, S. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.)