

Patience is a virtue: Individual differences in cue-evoked pupil responses under temporal certainty

Audrey V. B. Hood, Katherine M. Hart, Frank M. Marchak, Keith A. Hutchison

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.3758/s13414-022-02482-7>

Patience is a virtue: Individual differences in cue-evoked pupil responses under temporal
certainty

Audrey V. B. Hood, Katherine M. Hart, Frank M. Marchak, and Keith A. Hutchison

Montana State University

Please Address Correspondence to:

Keith A. Hutchison
Department of Psychology
Montana State University
P.O. Box 173440
Bozeman, MT 59717-3440 USA

E-mail: khutch@montana.edu
Phone: 406.994.5528

Abstract

Attention control is often examined behaviorally by measuring task performance and self-reported mind wandering. However, recent studies have also used pupillometry to measure task engagement versus task disengagement/mind wandering. In the current study, we investigated participants' ability to engage versus relax attention control *in anticipation* of hard (antisaccade) versus easy (prosaccade) trials within a saccade task, creating a "Cue-Evoked" Pupillary Response (CEPR). Participants completed the Automated OSPAN as a measure of working memory capacity (WMC) followed by a saccade task with a constant 5000 ms delay between cue and stimulus. Occasional thought-probes were included to gauge on- versus off-task attentional state. Consistent with recent findings (Hutchison et al., 2020; Wang et al., 2015), we found better performance and more Task-Unrelated Thoughts (TUTs) on prosaccade trials, larger pupil diameters when preparing for antisaccade trials, and larger pupil diameters when on-task. Further, lower WMC individuals showed pupil dilation throughout the fixation delay for both types of trials, whereas higher WMC individuals only showed dilation immediately before stimulus onset when expecting an antisaccade trial. Saccade accuracy was predicted by WMC, smaller early CEPR, larger late CEPR, and less CEPR variability, but not self-reported TUTs. These findings demonstrate that, under temporal certainty, higher WMC individuals may be more efficient in their exertion of attention control. Further, they indicate that physiological measures can not only validate self-report measures, but also help identify situations in which self-report may be.

Keywords: cognitive control, pupillometry, mind wandering, attention

Patience is a virtue: Individual differences in cue-evoked pupil responses under temporal certainty

Attention control refers to the ability to maintain focus in the face of distractions that can stem from either external (e.g., noises in the room) or internal (e.g., mind wandering) sources. This ability is often examined behaviorally by measuring task performance (e.g., Engle, 2002) and/or self-reported mind wandering (e.g., McVay & Kane, 2009). In addition to these measures, researchers have also examined physiological indices of attention. One physiological measurement that has recently received a great deal of renewed interest is pupillometry, as changes in pupil diameter in response to task demands provide insights into the functioning of the Locus-Coeruleus Norepinephrine-System (LC-NE), a system that plays a major role in attention preparation, mind wandering, and task performance.

Locus-Coeruleus-Norepinephrine System

The LC-NE system plays an integral role in controlling task engagement through modulating arousal, attention, and alertness (Gilzenrat et al., 2010; Unsworth & Robison, 2017a). The locus coeruleus (LC), located in the dorsal pons, is a neuromodulatory nucleus with projections throughout the neocortex and is responsible for most norepinephrine (NE) released in the brain. Although the LC has widespread projections, it demonstrates regional specificity, with brain areas involved in attentional processing (e.g., parietal cortex) and motor responding (e.g., pre-motor cortex) receiving especially dense LC-NE innervation (Foote & Morrison, 1987).

According to the Adaptive Gain Theory (Aston-Jones & Cohen, 2005; Aston-Jones et al., 2007), the LC-NE system is sensitive to current task utility (i.e., the likelihood that effortful responding will bring about task-related rewards). When task utility is high, the LC-NE system enables active task goal maintenance and utilization, increasing performance (i.e., exploitation).

Conversely, when task utility is low, it triggers task disengagement (i.e., exploration). As long as baseline arousal is above minimal levels, such exploitation versus exploration is expressed through two modes of firing: tonic and phasic (Usher et al., 1999). Under tonic mode, baseline LC activity is elevated and there is little-to-no phasic task-evoked response, reflecting disengagement from the current task and enhanced processing of task-unrelated stimuli. It is under this mode where participants may begin to mind wander and show declines in task performance. In contrast, under phasic mode, baseline LC activity is lower and phasic task-evoked responses are greater, producing increased NE release throughout the cortex as task demands increase, which increases the gain in processing task-relevant stimuli. This enhances the sensitivity of neurons within frontal-parietal regions responsible for maintaining, updating, and implementing task goals and for suppressing default-mode network (DMN) areas active during rest periods and internal thought (Raichle et al., 2001; Unsworth & Robison, 2017a). Thus, in an ideal situation, the LC modulates frontal-parietal regions so that attention can be fully engaged and allocated to goal-relevant stimuli. In typical attention control tasks, this leads to suppression of the DMN (i.e., task-unrelated thoughts), active maintenance of the current goal in working memory, and a strong phasic LC response that is coupled with current task demands, resulting in optimal performance.

Recently, Unsworth and Robison (2017a) have used the adaptive-gain theory to help explain poor performance in attention control tasks by individuals lower in working memory capacity (WMC). Specifically, they proposed that dysregulated arousal (i.e., inconsistent and inappropriate NE release) could impair functioning of critical neural networks, such as the fronto-parietal control network. This, in turn, could lead to increased activity in the DMN,

reflecting more mind wandering and ultimately poor task performance within these individuals.

Using Pupillometry to Measure LC-NE Activity

It is challenging to measure LC activity directly, although there have been recent advances using neuromelanin-sensitive magnetic resonance (Bachman et al., 2021; Keren et al., 2015; Ohtsuka et al., 2013; Sun et al., 2020). However, LC activity can be indirectly measured through pupillometry. Numerous studies have indeed demonstrated that pupil diameter closely tracks LC activity, serving as an index of phasic versus tonic LC modes of task engagement versus disengagement, respectively (Aston-Jones & Cohen, 2005; Franklin et al., 2013; Gilzenrat et al., 2010; Rajkowski et al., 1993; Reimer et al., 2016; Unsworth et al., 2018; Unsworth & Robison, 2016) [Although see Megemont et al., 2022 for evidence that pupil diameter may only explain a small portion of the variance in LC activity.]

Pupillometry and Task Engagement

Task-evoked pupillary responses (TEPRs) are changes in pupil diameter that coincide with changing task difficulty. In terms of task engagement, TEPRs have been obtained in a variety of tasks, including tasks of short-term memory (STM; Kahneman & Beatty, 1966), sustained attention (van den Brink et al., 2016), working memory (Kahneman & Beatty, 1966), cognitive control (Rondeel et al., 2015), and complex reasoning (Bradshaw, 1968; Hess & Polt, 1964). TEPRs are also a sensitive measure of mental effort, as pupils dilate during increased cognitive load and constrict as the load lightens (Beatty, 1982; Heitz et al., 2008; Hess & Polt, 1964; Kahneman, 1973; Peavler, 1974). Thus, pupil changes can provide an online index of the degree of attentional resources allocated to a task (van Der Meer et al., 2010). For instance, changes in pupil diameter are greater when processing difficult, as opposed to simple, sentences

(Just & Carpenter, 1993) and can reflect context maintenance and response preparation within the AX-CPT task (Chatham et al., 2009; Chiew & Braver, 2013).

Pupillometry and Task Disengagement

In terms of task disengagement, TEPRs are largely absent when people report mind wandering (Hutchison et al., 2020; Smallwood et al., 2011; Unsworth & Robison, 2017). During such attentional lapses, the pupillary response no longer tracks task difficulty, just as peoples' attention is likewise decoupled from the necessary task set. Similarly, frequent attentional lapses, as indicated by trial-to-trial variability in either baseline pupil diameter or in TEPRs, have been shown to correlate negatively with WMC (Aminihajibashi et al., 2020; Robison & Brewer, 2020; Robison & Unsworth, 2019; Unsworth, & Robison, 2015; 2017b), attention control (Madore et al., 2020; Unsworth & Robison, 2017b), and long-term memory (Madore et al., 2020).

Of interest to the current study, several studies have examined links between TEPRs and reports of mind wandering during cognitive tasks. For instance, Franklin et al. (2013) measured pupil diameter during reading and found a tonic disengagement pattern immediately before probes in which participants reported mind wandering. In addition, trial-to-trial variability in TEPRs correlates with more self-reported instances of mind wandering (Robison & Brewer, 2020; Unsworth & Robison, 2017b). TEPRs not only distinguish between on- and off-task states, but also between different types of off-task states (i.e., mind wandering vs. being distracted), suggesting that TEPRs can measure distinct types of attentional lapses (Unsworth & Robison, 2016a). Such findings led Unsworth et al. (2018) to conclude that “pupillary responses provide a consistent means of tracking fluctuations in intrinsic alertness and attention (linked to LC-NE and cortical sustained attention network functioning) during tasks that demand a great deal of sustained attention for optimal performance” (p. 1251). Thus, according to Unsworth and

colleagues, lower WMC individuals may not differ so much in their ability to exert attention control per se, but in their ability to do so consistently across trials, as they suffer from frequent attentional lapses.

The current study examines participants' ability to modulate attention control *in anticipation* of hard versus easy trials within a task, rather than during such trials themselves. In doing so, we revisit recent work from Hutchison et al. (2020) to address both a limitation and some intriguing findings that diverged from previous work. In the sections to follow, we provide a brief overview of Hutchison et al.'s methodology and results that are necessary for understanding the current study.

Cue-Evoked Pupillary Responses (CEPRs)

Cuing participants on the nature of each upcoming trial is an effective method for demonstrating the flexible engagement of cognitive control (Bugg & Smallwood, 2016; Hutchison, 2007; Hutchison et al., 2016). Because of this, recent studies have used pupillometry not just to measure TEPRs to ongoing task demands, but to measure participants' modulation of attention control *in anticipation* of easy or difficult trials within a task, creating what we will call a "Cue-Evoked" Pupillary Response (CEPR). Across multiple studies, individuals' CEPRs differ when anticipating difficult versus easy trials, indicating enhancement or relaxation of top-down control, depending upon the difficulty of the anticipated trial (Hutchison et al., 2020; Irons et al., 2017; Wang et al., 2015). Such findings demonstrate that CEPRs can accurately reflect effort in task-set preparation.

Recently, Hutchison et al. (2020) measured pupil diameter while participants completed a saccade task in which a cue preceded each trial instructing participants to look "toward" (prosaccade) or "away" (antisaccade) from the upcoming saccade stimulus to catch the target.

They also included occasional thought probes to measure self-reported mind wandering (Task-Unrelated Thoughts; TUTs) and to examine CEPRs separately when participants were task-focused versus mind wandering. Following work showing that individual differences in WMC correlate more strongly with antisaccade performance at longer delays (Meier et al., 2018; Moffitt, 2013), Hutchison et al. varied the post-cue fixation delay (500 – 8000ms) preceding the saccade stimulus to examine both attention preparation (i.e., task engagement) and mind wandering (i.e., task disengagement). In two experiments, participants demonstrated more positive CEPRs preceding antisaccade trials (as opposed to prosaccade) and preceding accurate responses (as opposed to errors), showing CEPRs are valid indicators of task-set preparation.

Interestingly, examining CEPRs separately when participants reported being on- or off-task (e.g., mind wandering), Hutchison et al. (2020) found that CEPRs differed across trial types only when participants reported being on-task, but not when mind wandering. Specifically, when on-task, CEPRs reflected more constricting of pupils (indicating relaxing control) when expecting a prosaccade trial, relative to an antisaccade trial. In contrast, when off-task, pupil diameters constricted equally over the delay regardless of trial type. Such results indicate that, when attention is on-task, trial type differences in CEPRs reflect the degree of preparatory control exerted for the upcoming trial. When attention is off-task, LC-NE activity, as measured by CEPRs, is decoupled from the current task, consistent with Smallwood and colleagues (2011; 2013) “decoupling hypothesis” and Unsworth and Robison’s (2017a) finding that off-task thoughts are associated with reduced phasic pupil responses. Hutchison et al. noted that the similarity in CEPRs when actively anticipating a prosaccade trial and during mind wandering suggests that preparing for prosaccade trials involves purposefully relaxing control, allowing reflexive saccades to be “captured” by the exogenous saccade stimulus. These findings indicate

different CEPR patterns across trial types do not simply reflect task-engagement versus disengagement per se. Rather; they represent the degree of preparatory effort required if engaged in the task such that, when on-task, participants can engage control when they need it (antisaccade trials) or relax control when they don't (prosaccade trials).

Hutchison et al. (2020) also posited that individual's higher in WMC may be more efficient in exerting or withholding such attention control. Indeed, there is evidence to suggest they may be more calibrated in terms of discriminating situations that demand exertion of cognitive control from situations that instead allow reliance on habitual responding. For instance, individuals higher in WMC are better able to detect errors and strategically adjust control to match current task goals (Coleman et al., 2015). Further, higher WMC individuals are more flexible in controlling when they engage in mind wandering (Rummel & Boywitt, 2014). Specifically, they are less likely to mind wander when the task requires concentration and effort (Kane et al., 2007; McVay & Kane, 2009) and more likely to mind wander when the task is easy (Kane et al., 2007; Levinson et al., 2012).

Questions and Limitations of Hutchison et al. (2020)

Although the Hutchison et al. (2020) findings bridged current theories regarding LC-NE functioning, task performance, and mind wandering, some findings potentially conflict with past research. In addition, there was a limitation that prevented us from examining WMC differences in performance. We discuss these issues below.

Variable Versus Fixed Delay

During the review process, an anonymous reviewer pointed out that Hutchison et al.'s (2020) CEPR pattern diverged from previous work (Wang et al., 2015) in that our pattern primarily reflected pupil constriction over time, rather than dilation. Specifically, when

participants were expecting prosaccade trials, their pupils monotonically constricted relative to baseline. In contrast, when participants were expecting antisaccade trials, their pupils either slightly dilated or remained flat (depending upon accuracy) during the first four seconds and then gradually constricted thereafter, suggesting an early engagement of cognitive effort followed by a steady decrease. In contrast, Wang et al. (2015) found initial constriction for both trial types combined with greater dilation for antisaccade than prosaccade trials in the 200 ms immediately preceding onset of the saccade stimulus.

In response to the reviewer's query, we hypothesized the different patterns may have resulted from the difference in temporal certainty between the two studies. Specifically, we used variable delays, whereas Wang et al. (2015) used a constant delay. Relatedly, Unsworth et al. (2018) also found worse performance and differential phasic pupillary responses when using a varied versus a fixed inter-stimulus interval. This difference is likely due to increased difficulty under variable delay, as participants must rapidly engage preparatory control and then maintain vigilance throughout the entire delay period. In contrast, a constant delay period requires less focused attention, allowing for increased control only immediately prior to the stimulus. As such, a constant delay typically results in better overall performance on sustained attention tasks (for discussion, see Unsworth et al., 2018). Indeed, when studies employ a constant interval, pupil dilation typically peaks immediately before stimulus onset (Bradshaw, 1968, 1969; Jennings et al., 1998; Richer & Beatty, 1987; Richer et al., 1983; Unsworth et al., 2018; van der Molen et al., 1989). These physiological findings suggest that, when delays are held constant, participants can be more efficient in attention allocation, exerting control immediately before the stimulus onset (rather than maintaining vigilance throughout the entire delay). In the current study, we examine

whether some individuals, such as those higher in WMC, may be better at such efficient calibration of control.

Pupil Versus Self-Report

Another intriguing finding from Hutchison et al. (2020) is that pupil diameter seemed to track attention preparation better than self-reported mind wandering. As expected, in Experiment 1, participant's pupils were considerably smaller in anticipation of prosaccade trials than antisaccade trials, suggesting participants allowed themselves to relax when expecting the habitual, prosaccade trials. Surprisingly, however, participants reported more mind wandering on antisaccade trials, which is inconsistent not only with their CEPRs, but with decades of past literature showing greater TUTs during easier tasks (Kane et al., 2007; Levinson et al., 2012). We theorized this discrepancy between the self-report measures and physiological data was due to the placement of thought probes within the saccade trials. Because the thought probes appeared after participants made a response, and because people performed much worse on antisaccade trials, it is possible that participants rationalized their attentional state as having been off-task during antisaccade trials. To test this, and to allow for a more accurate representation of participant's attentional state while preparing to make a saccade, in Experiment 2, we again presented the thought probe after the fixation delay, but in lieu of the saccade cue, rather than after the saccade cue and response. We also reduced the percentage of thought probes from 25% to 17% to reduce potential reminders to stay on-task caused by inserting frequent thought probes. Participants' pupil diameters were again considerably smaller in anticipation of prosaccade trials than antisaccade trials. Importantly, however, participants' TUT responses now matched their CEPRs (and past studies), with them reporting more mind wandering on prosaccade than antisaccade trials. These results not only highlight the validity of pupillometry as a measure of

attentional state, but also suggest that objective pupillometry measures can both validate self-report measures and help identify situations in which self-report may be inaccurate due to procedural problems.

WMC and Performance

Finally, a limitation to Hutchison et al. (2020) is that we were unable to report our OSPAN results due to data collection errors, resulting in too much missing data. Although the behavioral patterns replicated previous individual difference results, such that higher WMC individuals had greater antisaccade performance especially at longer delays (Meier et al., 2018; Moffitt, 2013), due to the high rate of missing data, we were unable to examine individual differences in WMC and how they relate to phasic pupil variability, TUTs, and performance.

Current Study

In the current study, we replicate Hutchison et al.'s (2020) study while adding three important changes. First, we used a constant 5000 ms delay, rather than their variable (500, 2000, 4000, or 8000 ms) delay. Second, we included the automated OSPAN (Unsworth et al., 2005) to examine CEPRs as a function of WMC as well as the relation between WMC and CEPRs, saccade accuracy, and mind wandering. Third, we simultaneously test for the contributions of WMC, pupil dynamics (i.e., CEPRs and CEPR variability across trials), and self-reported mind wandering in predicting saccade performance using multiple regression. This allowed us to examine the possibility that pupil diameters may track attention preparation better than self-reported TUTs in this paradigm.

Following previous work (Hutchison et al., 2020; Wang et al., 2015), we predicted greater accuracy and more TUTs on prosaccade trials, more positive CEPRs when preparing for antisaccade trials, and more positive CEPRs when on-task. Also, individuals higher in WMC

should have better saccade performance overall (Hutchison, 2007; Kane et al., 2001; Unsworth et al., 2004). Further, by using a fixed delay, we predict CEPR patterns will follow those in Wang et al. (2015), such that pupil dilation will be larger immediately before stimulus onset when expecting an antisaccade trial. Importantly, we expect this pattern primarily for individuals higher in WMC, as they may be more efficient at knowing when to exert control. Lastly, we predict that CEPR variability (indicating frequent attentional lapses) will predict saccade accuracy better than self-reported TUTs.

Method

Participants and Design

In accord with Simmons, Nelson, and Simonsohn (2011), we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. In terms of sample size, Hutchison et al. (2020) had usable data from 118 and 95 participants in their Experiments 1 and 2, respectively. Cohen's d for the interaction of Trial Type and Delay on four second trials was .76 in Experiment 1 and .78 in Experiment 2. We would only need 18 participants to achieve power of .95 to detect these interactions. However, because we also wanted to examine individual differences in accuracy, mind wandering, CEPRs, and pupil variability, we chose to run at least 120 participants and decided to continue running participants until the end of the semester, even if we had already reached our goal, to achieve enough participants after data exclusions. By the end of the semester, we had run 154 participants. Although we did not ask participants to report their age or gender, this population typically features freshman between 18-20 years old, of whom approximately 55-60% are female.

After data collection, we removed data from 19 participants because of technical issues with the eye tracker causing missing pupil data. This resulted in usable data from 135

participants, which gave us approximate power of .95 to obtain correlations of .3 or larger. Six additional participants were missing data from the AOSPAN task, so analyses that include WMC are based on 129 participants. Participants were tested individually in a laboratory session lasting approximately one hour. Prosaccade and antisaccade trials varied within subjects. We examined pupil diameter and target accuracy as a function of trial type, time, and self-reported mind wandering state.

Apparatus

We used the same equipment and experimental room for this study as we used for Hutchison et al. (2020). Specifically, we used E-studio E-prime software from Psychology Software Tools (version 2.0.8.90) to program and present the saccade stimuli and a Panasonic CF-50 ToughBook laptop, with a Mobile Intel Pentium 4-M 2.00 GHz processor, 768 MB of RAM, and an AT Mobility Radeon 7500 Display Adapter to run the experiment. We presented task stimuli on a 17-inch NEC Multisync LCD 1760v monitor, with 1024 x 768 screen resolution and a 60 Hz refresh rate, attached to the laptop via an RS232 USB serial port.

To measure pupil diameter, we used a contact-free, remote-controlled infrared eye camera (RED) with automatic gaze and head trackers designed by SensoMotoric Instruments (SMI). Participants could freely view the monitor without having to use a chinrest. The tracker had binocular temporal resolution of 120 Hz, with spatial resolution of 0.03° and gaze position accuracy at 0.4° . Participants sat approximately 60 cm from the RED camera positioned directly under the monitor presenting task stimuli. An RS232 USB serial port on the Panasonic ToughBook laptop allowed the SMI RED tracking software to communicate with the E-prime software that ran the saccade task.

Procedure and Stimuli.

This study received permission from the Institutional Review Board at Montana State University. Figure 1 displays the trial sequences for a normal and thought probe antisaccade trial. The procedure and stimuli were identical to Hutchison et al. (2020) except that each trial had a fixed duration of 5000 ms between the fixation and saccade cue. For normal trials, participants saw a light gray background that remained onscreen while the following stimuli were presented sequentially. All stimuli were presented in Courier New bold font. First, for 500 ms, either the word “toward” (in blue 18-point font) instructed participants to look toward an upcoming cue to catch a target (prosaccade trial) or the word “away” (in red 18-point font) instructed participants to look away from the cue to catch the target (antisaccade trial)¹. This was followed by a 1,000 ms blank screen. Next, a white 22-point central fixation cross (+) appeared and remained onscreen for 5000 ms (fixation period). Because we were primarily interested in CEPRs, the eye tracker recorded during the fixation screen only. Then, a 36-point white saccade cue (*) appeared on either the left or right side of the computer screen for 300 ms. Following this, either an ‘O’ or a ‘Q’ target in black 20-point font appeared on the opposite side as the cue for 100 ms and was immediately replaced (masked) by two ‘##’ symbols in black 25-point font, which remained onscreen for five seconds, or until target response. The cue, target, and mask appeared approximately 12.5 cm horizontally from the center of the fixation cross, resulting in approximately 11.89° visual angle between the location of the fixation cross and the location of the cue, target, and mask. Participants were instructed to identify the target by pressing either the ‘O’ or ‘Q’ button on the keyboard. Following a response, there was a 1000 ms inter-trial interval

¹ This trial cue was also 500 ms in Hutchison et al. (2020), although they misreported it as 1,500 ms.

preceding the next trial. As in Hutchison et al., luminance levels were 29 cd/m² for both the “away” and “toward” task cue screens and 30 cd/m² for all other trial screens.²

In addition to the normal trials, thought probes occurred on 13% of trials. Thought probe trials consisted only of the trial type cue and fixation period, followed immediately by the thought probe (see Figure 1). The thought probe question, based on McVay and Kane (2009), appeared in 14-point Courier New cyan font on a black background. The luminance level for this screen was 2 cd/m². Specifically, on thought-probe trials, participants saw the question “What were you just thinking about?,” along with seven response options: (1) *task* (i.e., thinking about the stimuli and the appropriate response); (2) *task performance* (i.e., evaluating one’s own performance); (3) *everyday stuff* (i.e., thinking about recent or impending life events or tasks); (4) *current state of being* (i.e., thinking about conditions such as hunger or sleepiness); (5) *personal worries* (i.e., thinking about concerns, troubles, or fears); (6) *daydreams* (i.e., having fantasies disconnected from reality); or (7) *other* (i.e., other thought types). [Following McVay and Kane (2009), we defined responses 3-7 as “off-task” thoughts.] Participants responded by pressing the corresponding number on the keyboard. Following the participant’s response, the next saccade trial began.

Participants first completed three practice blocks containing 12 trials each (36 total). The first practice block contained only prosaccade trials, the second block contained only antisaccade trials, and the final block contained six prosaccade trials and six antisaccade trials, presented in random order, designed to mimic the actual experiment. Participants were then instructed about

² This task relies on target-identification accuracy as an index of correct eye movements. As such, the timing of the cue, target, and mask are determined so that making an incorrect eye movement would lead one to miss the target presentation and must guess. We have indeed validated this task using eye movement measures. For instance, when participants made correct antisaccade eye movements, they were 90% accurate in correct target identification. In contrast, when participants incorrectly make a prosaccade on an antisaccade trial, their target identification was at chance (.50) (see Hutchison et al. 2020).

the thought probes. Following the practice blocks, participants completed two experimental blocks, with each block containing 60 prosaccade trials and 60 antisaccade trials, resulting in 240 total experimental trials (120 trials per block). All trials occurred in random order. The number of thought probes remained equal across blocks and saccade type, such that both blocks of 120 trials contained 104 normal trials and 16 thought probe trials (8 prosaccade, 8 antisaccade). The entire experimental session lasted approximately 1 hour.³

Prior to the saccade task, Participants first completed the Automated Operation Span (AOSPAN; Unsworth et al., 2005). During this task, participants were asked to solve simple math problems (e.g., $2 + 3 / 2 = ?$) while remembering letters in-between each math problem. After participants made a “true” or “false” decision via a mouse click on a math problem, a letter would appear for 250 ms for the participant to memorize. After each set of trials, a recall screen was presented listing 12 possible letters and the participant was instructed to click the mouse next to the letters in the correct order that they were presented. The task was composed of three blocks, with each containing 5 sets of between 3-to-7 trials, for a total of 75 letters and 75 math problems. The AOSPAN was scored by summing the total number of letters recalled in the correct serial position, as recommended by Conway and colleagues (Conway et al., 2005).⁴

Phasic Pupil Diameter Measurement

³ In addition to this task, participants completed the ASRS, a short measure of self-reported ADHD symptoms (Kessler et al., 2005) for exploratory purposes only. Aside from a positive correlation with antisaccade TUT rates ($r = .192, p = .026$), this measure did not correlate with any measure of saccade performance (all p 's $> .13$) and is not considered further. Nonetheless, we included it in the uploaded data file for anyone interested.

⁴ We did not use any exclusion criteria for AOSPAN performance. The positive relationship between processing accuracy and storage/recall suggests that using a processing score cutoff would remove more lower span than higher span individuals (see Richmond et al., 2021; Unsworth et al., 2009). Of note, our results remain consistent when an 85% criterion is used.

We measured pupillometry during the 5000 ms fixation period to examine the time course of CEPRs as a function of expected trial type. We excluded blink trials (in which pupil diameter measured zero) and trials in which the eye tracker failed to capture at least half of the possible observations (sampled approximately every 8 ms). These criteria removed an average of 9.74 trials (4.06%) per participant. For each trial, the first 30 ms of the fixation screen served as a baseline. We calculated CEPRs (averaged across eyes) for each 1-second bin by subtracting the 30 ms baseline from the average pupil diameter during that bin so that positive values reflect dilation and negative values reflect constriction. Complete data files can be found on the Open Science Framework at the following URL: <https://osf.io/gqxkm/>.

Results

For our analyses, we first examined the behavioral results of saccade target accuracy and thought-probe responses. In these analyses, we examine the overall effects and then enter AOSPAN scores in an ANCOVA to test interactions between task variables and WMC. Next, we examined the pupil diameter analyses of CEPRs and their relation to accuracy, WMC, and mind wandering. Finally, we conducted multiple regression analyses to identify unique contributions of WMC, CEPRs, CEPR variability, and self-reported mind wandering in predicting saccade performance. In all analyses, we use a two-tailed p -value of .05 as our criterion for significance.

Behavioral Results

Saccade Target Accuracy

Participants had higher accuracy on prosaccade trials ($M = .916$, $SE = .008$) than on antisaccade trials ($M = .650$, $SE = .010$), $t(134) = 25.555$, $p < .001$. When including WMC, there was a main effect of WMC [$F(1, 127) = 13.045$, $p < .001$, $\eta_p^2 = .093$], with overall better

performance among those with higher AOSPAN scores. Finally, the Trial Type x WMC interaction was significant [$F(1, 127) = 4.269, p = .041, \eta_p^2 = .033$] such that AOSPAN scores correlated more strongly with antisaccade performance ($r = .323, p < .001$) than prosaccade performance ($r = .170, p = .054$).

Thought-Probe Responses

Replicating Hutchison et al.'s (2020) Experiment 2, participants reported more TUTs on prosaccade trials ($M = .459, SE = .025$) than on antisaccade trials ($M = .431, SE = .023$), $t(134) = 2.069, p = .040$. When including WMC, neither the main effect of WMC [$F(1, 127) = 0.097, p = .757, \eta_p^2 < .001$] nor the Trial Type \times WMC interaction [$F(1, 127) = 0.170, p = .681, \eta_p^2 = .001$] were significant.

Pupil Diameter Analyses

Due to unequal variance across time periods, we corrected all such p -values using the Greenhouse-Geisser correction.

Cue-Evoked Pupil Response (CEPR)

During the 30 ms baseline period, pupil diameter was $.027 \pm .015$ mm ($\pm = 95\%$ confidence interval) larger for prosaccade trials ($M = 4.423$ mm) than antisaccade trials ($M = 4.396$ mm). This finding replicates Hutchison et al. (2020, Experiment 2) and suggests that participants shift to a tonic disengagement mode following the prosaccade cue (Usher et al., 1999). This disengagement mode is characterized by higher initial pupils and reduced CEPRs as LC activity becomes decoupled from the current task (Gilzenrat et al., 2010; Smallwood et al., 2011).

As in previous studies, we centered the data on baseline pupil diameter to get a clearer view of phasic pupil changes during the fixation period. Figure 2 shows the time course of

participants' CEPRs during the 5000 ms fixation period as a function of trial type. We used a 2 (Trial Type) \times 5 (Time) ANOVA to examine the time course of CEPRs during the 5000 ms fixation period. This analysis included CEPRs preceding accurate trials only. Overall, pupil diameter increased across time, $F(4, 536) = 8.751, p < .001, \eta_p^2 = .061$. However, there was a significant Trial Type \times Time interaction, $F(4, 536) = 24.194, p < .001, \eta_p^2 = .153$. Separate analyses conducted by Trial Type showed that, when preparing for an antisaccade trial, pupil diameter significantly increased across the fixation period [$F(4, 536) = 18.456, p < .001, \eta_p^2 = .121$] and this increase had both linear (overall increase) [$F(1, 134) = 22.850, p < .001, \eta_p^2 = .146$] and cubic (increase, flat, increase) [$F(1, 134) = 7.624, p = .006, \eta_p^2 = .054$] trends. However, when anticipating a prosaccade trial, pupil diameter remained flat across the fixation period, $F(4, 536) = 2.002, p = .147, \eta_p^2 = .015$. This pattern of pupil dilation that is greater when anticipating an antisaccade response replicates the pattern obtained by Wang et al. (2015) under conditions of temporal certainty, but does not replicate the more flattened or constricted pupil response pattern obtained by Hutchison et al. (2020) under temporally uncertain conditions.

CEPR and Accuracy

The next two analyses examined how the CEPR pattern differed as a function of response accuracy. We first examined CEPRs separately for antisaccade trials preceding accurate versus inaccurate responses. Although this analysis gives us a general idea of successful versus unsuccessful CEPR patterns, its disadvantages are that (1) chance performance is 50% in the task, which adds noise to these CEPR patterns, and (2) we could only examine antisaccade performance due to ceiling effects on prosaccade trials, with 24 participants having accuracy at or above 98%. Therefore, in a second analysis, we examined antisaccade and prosaccade CEPRs

separately using individual differences in response accuracy as a covariate to see how CEPRs differed as a function of participant accuracy.

To examine CEPRs and accuracy on antisaccade trials, we used a 2 (Accuracy) \times 5 (Time) ANOVA to examine CEPRs preceding accurate and erroneous responses. These data are shown in Figure 3. There was no overall effect of Accuracy, $F(1, 134) = 2.357, p = .127, \eta_p^2 = .017$. However, there was a significant effect of Time [$F(4, 536) = 14.959, p < .001, \eta_p^2 = .100$], with overall pupil dilation across the fixation period. Moreover, there was a Time \times Accuracy interaction [$F(4, 536) = 3.798, p = .014, \eta_p^2 = .028$] showing that pupil dilation across the fixation period was greater preceding accurate responses than preceding errors. To test this interaction, we separately examined changes in pupil diameter during the first two seconds (0-2 seconds) and last two seconds (3-5 seconds) of the fixation period. This analysis confirmed the observation above. For the first two seconds, there was only a main effect of Time [$F(2, 268) = 6.247, p = .009, \eta_p^2 = .045$] and a marginal main effect of Accuracy [$F(1, 134) = 3.687, p = .057, \eta_p^2 = .027$], but no Time \times Accuracy interaction, $F(2, 268) = 0.822, p = .408, \eta_p^2 = .006$. In contrast, for the last two seconds, there was both a main effect of Time [$F(2, 268) = 19.823, p < .001, \eta_p^2 = .129$] and a Time \times Accuracy interaction [$F(2, 268) = 7.435, p = .003, \eta_p^2 = .053$], with steeper pupil dilation preceding correct responses (see Figure 3).

During the 30 ms baseline period, pupil diameter was $.018 \pm .015$ mm ($\pm = 95\%$ confidence interval) larger preceding error trials ($M = 4.414$ mm) than correct trials ($M = 4.396$ mm). This finding is consistent with the Trial Type effect observed above in that trials in which participants were disengaged are characterized by larger baseline pupils and reduced CEPRs. Further, this finding replicates Gilzenrat et al. (2010), who observed larger baseline pupils preceding trials associated with errors or extremely long RTs. [Although not reported in the

paper, a reanalysis of Hutchison et al. (2020) shows that baseline pupils were numerically larger preceding errors than accurate responses; however, this difference did not reach significance in either experiment ($p = .42$ and $.22$ in Experiments 1 & 2, respectively).]

We next examined how CEPR patterns differed as a function of participant accuracy by analyzing CEPRs for each trial type using the full range of individual differences in accuracy for that trial type as a covariate. In both the antisaccade and prosaccade analyses, we therefore used a Time (5 seconds) x Participant Accuracy (continuous) ANCOVA to examine CEPRs as a function of participant accuracy. For the antisaccade trial analysis, there was a Participant Accuracy \times Time interaction [$F(4, 532) = 8.055, p = .002, \eta_p^2 = .057$] driven by a quadratic pattern [$F(1, 133) = 39.564, p < .001, \eta_p^2 = .229$] such that individuals lower in overall accuracy showed earlier pupil dilation whereas individuals higher in overall accuracy showed late pupil dilation. In contrast, for the prosaccade trial analysis, there was only a main effect of Participant Accuracy, $F(1, 133) = 7.163, p < .008, \eta_p^2 = .051$. Specifically, participants lower in accuracy showed increased dilation throughout the fixation period. There were no correlations between individual differences in accuracy and baseline pupil diameter on either the antisaccade trials ($r = -.100, p = .249$) or prosaccade trials ($r = -.092, p = .288$).

In order to illustrate these ANCOVA findings, we next split the participants into tertiles based on their accuracy on each trial type to show separate CEPR patterns for “good performers” (top 33rd percent), “middle performers” (middle 33rd percent) and “poor performers” (bottom 33rd percent). This corresponded to mean scores (range in parentheses) for good, middle, and poor performers, respectively, of .78 (.71-.91), .64 (.58-.71), and .53 (.39-.58) on antisaccade trials and .97 (.96-1.00), .94 (.91-.96), and .82 (.47-.91) on prosaccade trials. Figure 4 shows CEPR patterns for the good performers (solid line), middle performers (dashed line), and poor

performers (dotted line) for antisaccade trials (top graph) and prosaccade trials (bottom graph). Visual inspection of Figure 4 confirms the observations from the ANCOVA above. Specifically, poorer performers had early, and steady, pupil dilation that peaked at around 4 seconds, whereas good performers had no dilation until the last couple seconds (for antisaccade) and peaked right before onset of the saccade cue.

To further test these novel observations, we examined pupil diameter changes across time points for each tertile group separately. For antisaccade trials, good performers showed increases in pupil diameter only during seconds three-to-four [$t(45) = 3.561, p = .001$] and four-to-five, $t(45) = 7.030, p < .001$. In contrast, poor performers' pupil diameter increased from zero-to-one second [$t(43) = 5.304, p < .001$] and again from two-to-three seconds, $t(45) = 2.813, p = .007$. Middle performers showed a mix of both patterns, with significant increase at zero-to-one [$t(44) = 3.512, p = .001$], two-to-three [$t(44) = 2.218, p = .032$], and four-to-five seconds, $t(44) = 2.945, p = .005$. For prosaccade trials, poor performers again showed early pupil dilation that was significant from zero-to-one seconds [$t(39) = 4.426, p < .001$] and two-to-three seconds, $t(39) = 2.511, p = .016$. In contrast, both good and middle performers showed significant pupil constriction from one-to-two seconds [$t(50) = 2.652, p = .011$; $t(43) = 2.271, p = .028$ for good and middle performers, respectively] and no changes thereafter.

CEPR and WMC

Next, we examined how CEPRs differed across individual differences in WMC. During the initial 30 ms baseline, pupil diameter was unrelated to WMC preceding either antisaccade ($r = -.059, p = .508$) or prosaccade ($r = -.044, p = .618$) trials. As was the case for participant accuracy, we first examined individual differences in CEPRs when preparing for antisaccade or prosaccade trials using a 2 (Trial Type) \times 5 (Time) \times WMC continuous ANCOVA. There was a

significant Time \times WMC interaction [$F(5, 635) = 4.557, p = .019, \eta_p^2 = .035$] that showed significant linear [$F(1, 134) = 21.552, p < .001, \eta_p^2 = .139$] and quadratic [$F(1, 134) = 14.263, p < .001, \eta_p^2 = .096$] trends. The linear pattern reflects greater pupil dilation across the fixation period for individuals lower in WMC. The quadratic pattern shows that this difference in dilation is in the first three seconds only. The 3-way WMC \times Trial Type \times Time interaction was not significant, $F(4, 508) = 1.505, p = .225, \eta_p^2 = .012$.

For illustrative purposes, we next split the participants into tertiles based on their AOSPAN scores to show separate CEPR patterns for “High WMC” (top 33 percent), “Middle WMC” (middle 33 percent) and “Low WMC” (bottom 33 percent) individuals. This corresponded to mean AOSPAN scores (range in parentheses) for high, middle, and low WMC, respectively, of 58 (46-75), 37 (30-45), and 19 (6-29). Figure 5 shows CEPR patterns for the High WMC (solid line), Middle WMC (dashed line), and Low WMC (dotted line) individuals on antisaccade trials (top graph) and prosaccade trials (bottom graph). As can be seen in Figure 5, individuals lower in WMC showed pupil dilation throughout the fixation period for both trial types, whereas those higher in WMC showed flattened pupil responses throughout most of the fixation period, with an increase only at the end when expecting an antisaccade trial.

To further test these novel observations, we examined CEPRs for each group separately. Collapsed across trial types, low WMC individuals had significant increases in pupil diameter across each of the first three seconds [$t(44) = 4.946, p < .001, t(44) = 2.796, p = .008, t(44) = 3.472, p = .001$ for pupil increases across 0-1, 1-2, and 2-3 seconds, respectively]. In contrast, high WMC showed an increase in pupil diameter only during the final second [$t(41) = 2.159, p = .037$] that was significant when preparing for antisaccade [$t(41) = 3.114, p = .003$], but not prosaccade trials, $t(41) = 0.560, p = .579$. Middle WMC individuals showed overall significant

constriction from one-to-two seconds [$t(41) = 2.590, p = .013$] and dilation from four-to-five seconds, $t(41) = 2.208, p = .033$.

Summarizing antisaccade performance across Figures 2-5, one can see that the overall increasing pupil dilation for antisaccade trials (Figure 2) is due to early dilation among lower WMC individuals (and poor performers generally) and late dilation for higher WMC individuals (and good performers generally). Similarly, the flat overall pupil diameter for prosaccade trials is driven by opposite effects in that there was pupil dilation among those lower in WMC, but constriction among those with at least medium WMC.

CEPR and Mind Wandering

We next examined pupil diameter separately for trials in which participants reported being on-task versus off-task. This analysis only included the 13% of trials in which participants received a thought probe and only included participants who reported both TUT and task-related thought responses for each trial type. Because of this, data were missing from 21 participants (remaining $N = 114$). To avoid removing additional participants, we coded thought probe responses of both 1 (thinking about task) and 2 (thinking about task performance) as “on-task,” as in Hutchison et al. (2020).

We used a 2 (Trial Type) \times 2 (Probe Response) \times 5 (Time) ANOVA to examine CEPRs as a function of Trial Type and Probe Response. There was no effect of Probe Response [$F(1, 113) = 0.00, p = .993, \eta_p^2 = .000$], and Probe Response did not interact with Time [$F(4, 452) = 0.400, p = .782, \eta_p^2 = .004$], or with Trial Type, $F(1, 113) = 0.149, p = .700, \eta_p^2 = .001$.

Similarly, when WMC was added as a covariate, none of the effects involving Probe Response interacted with WMC (all F 's $< 2.34, p$'s $> .130, \eta_p^2$'s $< .022$).

In addition, we tested for baseline pupil diameter differences as a function of Probe Response and Trial Type using a 2 (Trial Type) \times 2 (Probe Response) ANOVA. No effects were significant (all F 's < 2.270 , p 's $> .135$, η_p^2 's $< .020$).

Predicting Saccade Performance

Intercorrelations

Prior to predicting individual differences in saccade performance, we examined intercorrelations between saccade accuracy, AOSPAN, phasic pupil response (CEPR, with positive values indicating dilation), variability in phasic pupil response across trials (CEPR SD), and TUT rates. Because the earlier Accuracy \times Time interactions showed that the first two seconds and the last two seconds of the fixation period were differentially related to performance, we examined the CEPR separately for these two time periods. These correlations are shown in Table 1. As shown in Table 1, antisaccade and prosaccade accuracy correlated similarly with the predictor variables, although the correlations were stronger for antisaccade accuracy. Consistent with our earlier analysis, saccade accuracy was negatively correlated with early CEPR and positively correlated with late CEPR. In addition, saccade accuracy was negatively correlated with CEPR variability across trials and with TUTs. However, both TUT rates and late CEPR correlated with antisaccade accuracy, but not prosaccade accuracy. In addition, early CEPR, late CEPR, and CEPR variability were highly intercorrelated across prosaccade and antisaccade trials (Pearson's r above .790). Finally, TUT rates did not correlate with early CEPR, but were negatively correlated with late CEPR.

Multiple Regression

We next predicted antisaccade accuracy, prosaccade accuracy, and combined accuracy based on individual differences in WMC, early CEPR, late CEPR, CEPR variability, and TUT

rate. In each case, the chosen predictors matched the criterion measure, such that we used measures calculated from antisaccade trials, prosaccade trials, and collapsed across trials to predict antisaccade, prosaccade, and overall accuracy, respectively. These analyses are shown in Table 2. When predicting antisaccade accuracy, all predictors were significant except for TUT rate ($\beta = -.109, t = 1.406, p = .146$). Specifically, antisaccade accuracy was higher for individuals high in WMC ($\beta = .257, t = 3.531, p = .001$), individuals with smaller early CEPR ($\beta = -.158, t = 2.122, p = .036$), individuals with larger late CEPR ($\beta = .356, t = 4.696, p < .001$), and individuals with less variable CEPRs across trials ($\beta = -.173, t = 2.222, p = .028$). Thus, although self-reported TUT rates were correlated with antisaccade accuracy, this novel finding indicated that they no longer predicted accuracy when entered together with the other predictors. When predicting prosaccade accuracy, the only significant predictors were early CEPR and CEPR variability. Specifically, prosaccade accuracy was higher for individuals with smaller early CEPR ($\beta = -.192, t = 2.138, p = .034$) and with less CEPR variability across prosaccade trials ($\beta = -.191, t = 2.041, p = .043$). Finally, when predicting overall accuracy, the pattern looked identical to predicting antisaccade accuracy, with higher accuracy for individuals higher in WMC ($\beta = .259, t = 3.397, p = .001$), with smaller early CEPR ($\beta = -.170, t = .159, p = .033$), with larger late CEPR ($\beta = .270, t = 3.418, p < .001$), and with less CEPR variability ($\beta = -.269, t = 3.272, p = .001$). As with antisaccade accuracy, WMC, early CEPR, late CEPR, and CEPR variability all predicted unique variance in accuracy. In contrast, self-reported TUT rates did not predict accuracy ($\beta = -.012, t = 0.143, p = .887$) when entered together with the other predictors.

For comparison, we performed a similar multiple regression predicting antisaccade accuracy, prosaccade accuracy, and combined accuracy based on CEPR, CEPR variability, and TUT rate using the combined data from Hutchison et al.'s Experiments 1 and 2 (2020, N = 206).

These analyses are presented in the supplementary section. To anticipate, CEPR variability was the only significant predictor in all three analyses. Specifically, CEPR variability was the only significant predictor when predicting antisaccade accuracy ($\beta = -.289, t = 5.370, p < .001$), prosaccade accuracy ($\beta = -.089, t = -2.525, p = .012$), and accuracy overall ($\beta = -.181, t = 5.085, p < .001$). As was found for the current data, although self-reported TUT rates were correlated with accuracy, ($r = -.148, p = .033$), they did not predict accuracy when entered together with the other predictors.

Discussion

We investigated participants' ability to engage versus relax attention control *in anticipation* of antisaccade versus prosaccade trials, creating a "Cue-Evoked" Pupillary Response (CEPR). The results of the current study both replicate and extend previous findings and provide important contributions to our understanding of the relation between WMC, pupil dynamics, self-reported mind wandering, and saccade performance. In terms of replication, consistent with Hutchison et al. (2020) and Wang et al. (2015), we found greater accuracy on prosaccade trials, more TUTs on prosaccade trials, larger baseline pupil diameters on prosaccade trials than antisaccade trials, larger CEPRs when preparing for antisaccade compared to prosaccade trials, and larger CEPRs when on-task vs off-task. Also, consistent with previous studies (Kane et al., 2001; Unsworth et al., 2004; Moffitt, 2013), individuals higher in WMC had better saccade performance, especially on antisaccade trials. Finally, as with Gilzenrat et al. (2010), baseline pupil diameters were larger preceding errors than preceding correct responses.

Importantly, the current study also revealed three novel findings. First, these results demonstrate that temporal certainty allows more efficient and delayed exertion of control, as revealed through CEPR patterns. Second, as predicted, such efficiency varied across individual

differences in WMC such that lower WMC individuals exerted control early and even for prosaccade trials, whereas higher WMC individuals efficiently exert control only in the last couple seconds when preparing for an antisaccade trial. Third, across multiple data sets, we demonstrated that pupil dynamics involving CEPR patterns and CEPR variability predicted accuracy, whereas self-reported TUTs did not. We expand upon each of these important novel contributions below.

Temporal Certainty Versus Uncertainty

One of the main goals of the current study was to examine CEPRs using a fixed delay in the saccade task to examine whether the difference in CEPR patterns in Hutchison et al. (2020) and Wang et al. (2015) were likely due to the two studies using variable versus fixed delays, respectively. Recall in Hutchison et al. (2020), when participants were expecting prosaccade trials, their pupils monotonically constricted relative to baseline. In contrast, when participants were expecting antisaccade trials, their pupils either slightly dilated or remained flat (depending upon accuracy) during the first four seconds and then gradually constricted thereafter, suggesting an early engagement of cognitive effort followed by a steady decrease. In contrast, Wang et al. (2015) found initial constriction for both trial types combined with greater dilation for antisaccade than prosaccade trials in the 200 ms immediately preceding onset of the saccade stimulus.

The fact that the current CEPR patterns follow Wang et al.'s (2015), who used a constant delay in a saccade task, as well as numerous studies that used constant delays in tasks other tasks (Bradshaw, 1968, 1969; Jennings et al., 1998; Richer & Beatty, 1987; Richer et al., 1983; Unsworth et al., 2018; van der Molen et al., 1989), supports the reasoning that the differences in CEPR patterns found between Hutchison et al. (2020) and the current study was likely due to the

variable versus fixed delays used in the two studies. This result provides further evidence that, when delays are held constant, participants can be more efficient in attentional allocation, exerting control immediately before the stimulus onset (rather than maintaining vigilance throughout the entire delay).

Good Versus Poor Performers & WMC

Importantly, in addition to showing temporal certainty allows greater efficiency in control, ours is the first study to demonstrate that it is only those higher in WMC that make use of this temporal certainty to efficiently withhold attentional effort until needed. Specifically, lower WMC individuals (and poor performers overall) had early and steady pupil dilation, whereas higher WMC individuals (and good performers generally) had no dilation until the last couple of seconds (for antisaccade trials) and peaked right before the onset of the saccade cue. These differences in CEPR patterns demonstrate how individual differences in WMC relate to the efficiency of effortful engagement. Under predictable timing, lower WMC individuals exert effort early, even for prosaccade trials, whereas higher WMC individuals efficiently exert effort only at the necessary time.

Further, these results provide evidence that the poorer performance from lower WMC individuals isn't due to them exerting less effort or control; instead, they are simply inefficient in the exertion of their control. Specifically, engaging control early on each trial requires considerable effort in maintaining such vigilance throughout the delay. These results are consistent with recent findings from Unsworth et al. (2020), who examined individual differences in preparatory activity during an interstimulus interval period within the psychological vigilance task. They found that individuals susceptible to lapses of attention (as measured by the slowest quintile of reaction times) demonstrated an increased pupillary

response, more pupil variability, and more TUTs during the interstimulus interval compared to individuals less susceptible to lapses of attention. These results suggest that individual differences in the ability to voluntarily control the intensity of attention (“intrinsic alertness”) may lead to lapses of attention and the ability to fully engage preparatory processes on a moment-to-moment basis. Unsworth et al. also found that, at the latent level, WMC predicted attention control, which in turn was related to off-task thoughts, increases in the pupillary response, and decreases in pupil variability during the interstimulus interval. The fact that WMC did not have a direct effect suggests the relation between WMC and lapses of attention was mediated by attention control abilities. Our results add to these findings, indicating how individual differences in WMC can relate to temporal dynamics of preparatory activity and the ability to efficiently engage control.

In regard to the overall literature, we believe these findings concerning poor efficiency have important implications for how researchers define and conceptualize individual differences in WMC. For instance, early studies examining individual differences in WMC assumed such differences were due to a differential amount of resources (Anderson et al., 1996; Case, 1972; Just & Carpenter, 1980; Ma et al., 2014). However, in the early 2000s, Engle and colleagues argued that WMC does not reflect a “capacity” or “amount” per se, but instead reflects the ability to control attention in the face of distraction (Engle, 2002; Kane & Engle, 2003). Strong evidence for this goal maintenance account of WMC performance differences comes from studies demonstrating individual differences in performance on conflict tasks such as Antisaccade and Stroop that involve maintaining only one item in working memory; the goal of the task itself (Engle, 2002; Kane et al., 2001). [See Hood & Hutchison (2021) for direct support of the goal maintenance account for WMC differences in the Stroop task]. More

recently, Unsworth (2015) argued that even the concept of “control” itself might be misleading. Specifically, individual differences in attention abilities are influenced not simply by the “amount” of control, but by the consistency with which control is engaged, such that lower WMC individuals might exert equal control on most trials, but suffer more often from attentional lapses than higher WMC individuals. Further, Unsworth and Miller (2021) suggest that both the consistency and intensity (strength) of attention vary between individuals and that this variation explains differences in task performance.

We can now add to these conceptualizations that WMC performance differences are due to not just the *consistency* of attention control, but also the *efficiency* in which that control is engaged. Thus, as a literature, we are moving away from theoretically empty descriptions of attention control involving capacity or resources and toward more operationally-definable terms of consistency and efficiency with which control is engaged.

Predicting Accuracy from Pupil Versus Self-Report Measures

Saccade accuracy was predicted by WMC, smaller early CEPR, larger late CEPR, and less CEPR variability, but not self-reported TUTs. Thus, although TUT rates had zero-order correlations with antisaccade accuracy, they no longer predicted accuracy when entered together with the other predictors. This same pattern was also found when reevaluating the Hutchison et al. (2020) data set, where TUT rates no longer predicted accuracy when entered together with CEPR and CEPR variability. Together, these findings indicate that physiological measures may provide a better indication of attentional state than self-report. This could be problematic, as the probe-caught method is commonly used to measure mind wandering. However, although thought-probes may catch instances of mind wandering before they reach awareness (Chin et al., 2012) and typically produce a good estimate of mind wandering frequency (Chin et al., 2012;

Smallwood & Schooler, 2006), there are important potential issues with this method. One issue is that self-report thought probes require that participants be consciously aware, and understand the content, of their thought processes during the ongoing task (Nisbett & Wilson, 1977; Smallwood & Schooler, 2006). Some individuals may have difficulty with this level of introspection, especially those low in conscious awareness of experiences and behaviors. An example of such “temporal dissociations” (Schooler, 2002) is when people fail to notice they have “zoned out” while reading. Because of these potential issues, cautious interpretation of self-reported mind wandering measures has been openly acknowledged (Schooler & Schreiber, 2004), and there have been compelling arguments questioning the credibility of these measures (Jack & Roepstorff, 2002; Jack & Shallice 2001; Lambie & Marcel, 2002). To further explore the credibility of TUTs, we next revisited an intriguing finding from Hutchison et al. (2020).

Revisiting Hutchison et al. (2020): Do Procedural Changes Affect TUT Rate?

As discussed in the introduction, an interesting finding by Hutchison et al. (2020) was that, despite pupillometry results suggesting more mind wandering on prosaccade trials for both experiments, participants reported more TUTs on antisaccade trials in Experiment 1 in which thought probes followed saccade responses, but more TUTs on prosaccade trials in Experiment 2 in which thought probes appeared in lieu of the saccade cue, rather than after the saccade response. Hutchison et al. (2020) stated this may have been due to participant reactivity by attributing their poor antisaccade performance to mind wandering. More recently, however, Kane et al., (2021) argued that if Hutchison et al.’s (2020) TUT reports for antisaccade trials were artificially inflated in Experiment 1, then the change in thought-probe procedures in Experiment 2 should have reduced their antisaccade TUT rates relative to Experiment 1. We think this is an important observation and worthy of further exploration.

Before providing explanations for the different TUT pattern across experiments, we first examined whether the pattern of TUT rates significantly differed across the three studies (Hutchison et al., 2020 Experiments 1 & 2 and the current study). This analysis is presented in the supplementary section. Figure 6 shows the TUT rates across the three studies as a function of Trial Type. Visual inspection of Figure 6 shows that the effect of Trial Type flipped when thought probes occurred in the absence of saccade responses, relative to Experiment 1 in which thought probes occurred after saccade responses. This observation is supported by the finding that TUT rates for prosaccade trials differed across experiments, but TUT rates for antisaccade trials stayed the same (see supplementary analysis).

Given the specific pattern in which prosaccade TUT rates differed across procedure whereas antisaccade TUTs remained flat, we now suggest that, in Experiment 1, participants were not reacting to antisaccade errors to assume mind wandering, but were instead reacting to prosaccade accurate target detection to assume task-focused attention. Given the nature of the task, and 50% chance accuracy, it makes sense that participants might be uncertain of their accuracy on antisaccade trials, but would nonetheless be certain of target detection on prosaccade trials. This phenomenological experience of accurate target detection could cause them to miss any mind wandering that might have occurred during the preparatory period. Asking the thought probe immediately after the preparatory period in Experiments 2 and 3 eliminated any such accuracy-based reactivity. We believe this second explanation parsimoniously explains the interaction pattern shown in Figure 6. It is important to stress that our finding of a Trial Type \times Experiment interaction disproves any hypothesis that TUT reports were validly reported across studies. Instead, these results demonstrate that the commonly used self-reported TUTs can be

prone to bias, especially when experimental methods allow for it, implicating cautious interpretation of self-reported mind wandering measures.

Limitations and Future Directions

There are a few limitations in the current study that could guide future research in this area. First, one of the main foci in the current study was to examine individual differences in task-set preparation using a fixed, rather than variable, preparatory period to understand the how WMC relates to both consistency and efficiency of attention control. As predicted, the basic CEPR results in this fixed paradigm mirrored that of Wang et al. (2015), rather than Hutchison et al. (2020). However, we recommend future studies employ a within-subjects manipulation to provide stronger evidence for delay predictability as driving the observed effects. [We did not originally see these projects as compatible, and so it did not occur to us to do this manipulation directly.] We also note that, in retrospect, this could potentially have been included as a third experiment in Hutchison et al. (2020). However, as previously explained, it was only during the Hutchison et al. (2020) review process that we were made aware of the discrepancy between the CEPR patterns in Hutchison et al. (2020) and Wang et al. (2015).

In addition, a second limitation to the current study is that we only used the AOSPAN as our measure of WMC. This is suboptimal because, although the AOSPAN measures WMC, it also measures factors unrelated to this ability such as speed of solving math problems (Foster et al., 2015; Loehlin, 2004; Wittman, 1988). Instead, multiple complex span tasks should ideally be administered to create a composite or factor score, consisting of the variance shared between the tasks (Conway et al., 2005). We chose to use only the AOSPAN task because it allowed us to keep the study length under one-hour. Further, this data was collected before the shortened span tasks were published (Foster et al., 2015). Future studies could use single blocks of multiple

complex span tasks, which could capture more variance attributed to WMC without adding much length to the study duration.

Third, because we only collected eyetracking data during the preparatory period, we were unable to measure saccade reaction time. Future studies could also engage the eyetracker during presentation of the saccade stimulus and target stimulus to examine how the temporal dynamics of preparatory control relate to faster saccade velocity after stimulus onset.

Finally, an important area for future work is to examine why self-reported TUTs appear to validly capture attentional state in some tasks, but not others. One consideration is the type of thought-probe used. For instance, Kane et al. (2021) recently found that content probes (asking participants *what* they were thinking about when mind wandering, as in the current study) are less susceptible to reactivity, confabulation, and bias than other types of thought probes, at least within the Sustained Attention to Response and Flanker tasks. However, it is also possible that the type of task in which thought probes are inserted may impact the validity of mind wandering reports (Kane et al., 2021). Our studies show the importance of eliminating task factors that can contribute to such bias. Given that thought probes are commonly used to measure instances of mind wandering and the importance of validly measuring attentional state, future studies should further explore the conditions under which self-reported TUTs validly capture attentional state.

Conclusion

Attention control is often measured at the behavioral level, such as by measuring task performance and/or self-reported mind wandering. In addition to these measures, researchers have also examined physiological indices of attention, such as pupillometry. In the current study, we investigated individual differences in the ability to engage versus relax attention control *in anticipation* of hard (antisaccade) versus easy (prosaccade) trials, with the inclusion of thought-

probes to measure self-reported mind wandering. The current results demonstrate that, under temporal certainty, higher WMC individuals are more *efficient*, engaging control only when anticipating difficult trials and waiting until immediately prior to the onset of such trials. Further, our results indicate that physiological measures can not only validate self-report measures, but also identify situations in which self-report may be inaccurate due to procedural problems.

Open Practices Statement

The experiment was not preregistered. Complete data files are available on the Open Science Framework at the following URL: <https://osf.io/gqxkm/>.

References

- Aminihajibashi, S., Hagen, T., Andreassen, O. A., Laeng, B., & Espeseth, T. (2020). The effects of cognitive abilities and task demands on tonic and phasic pupil sizes. *Biological Psychology, 156*, 107945.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 3*, 411-423.
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limits on retrieval. *Cognitive Psychology, 30*, 221-256.
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus–norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403-450.
- Aston-Jones, G. S., Iba, M., Clayton, E., Rajkowski, J., & Cohen, J. (2007). The locus coeruleus and regulation of behavioral flexibility and attention: Clinical implications. In G. A. Ordway, M. A. Schwartz, & A. Frazer (Eds.), *Brain norepinephrine: Neurobiology and therapeutics* (pp. 196-235). Cambridge: Cambridge University Press.
- Bachman, S. L., Dahl, R. J., Werkle-Bergner, M., Düzel, S., Forlim, C. G., Lindenberger, U., Kühn, S., & Mather, Mara. (2021). Locus coeruleus MRI contrast is associated with cortical thickness in older adults. *Neurobiology of Aging, 100*, 72-82.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276-292.
- Bradshaw, J. L. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology, 20*, 116-122.

- Bradshaw, J. L. (1969). Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science, 14*, 271-272.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. [Eds.] *Testing structural equation models*. Newbury Park, CA: Sage, 136-162.
- Bugg, J. M., & Smallwood, A. (2016). The next trial will be conflicting! Effects of explicit congruency pre-cues on cognitive control. *Psychological Research, 80*, 16-33.
- Byrne, B. M. (1998). *Multivariate applications book series. Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Lawrence Erlbaum Associates Publishers.
- Case, R. (1972). Validation of a neo-Piagetian mental capacity construct. *Journal of Experimental Child Psychology, 14*, 287-302.
- Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences, 106*, 5529-5533.
- Chiew, K. S., & Braver, T. S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology, 4*, 15.
- Chin, J., Mrazek, M., & Schooler, J. (2012). Blind spots to the self. In S. Vazire, & T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 79). New York: The Guildford Press.
- Coleman, J. R., Watson, J. M., & Strayer, D. L. (2018). Working memory capacity and task goals modulate error-related ERPs. *Psychophysiology, 55*, 1-14.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, Z. D., Wilhelm, O. & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786.

- Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL*. London: Sage Publications.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.
- Foote, S. L., & Morrison, J. H. (1987). Extrathalamic modulation of cortical function. *Annual Review of Neuroscience, 10*, 67-95.
- Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology, 66*, 2289-2294.
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & cognition, 43*, 226-236.
- Gilzenrat, M. S., Neuenhuis, S., Jepka, 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*, 252-269.
- Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R.W. (2008). Effects of incentive and working memory capacity: Behavioral and pupillometric data. *Psychophysiology, 45*, 119-129.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190-1192.
- Hood, A. V. B., & Hutchison, K. A. (2021). Providing goal reminders eliminates the relationship between working memory capacity and Stroop errors. *Attention, Perception, and Psychophysics, 83*, 85-96.

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 645-662.
- Hutchison, K. A., Bugg, J. M., Lim, Y. B., & Olsen, M. R. (2016). Congruency precues moderate item-specific proportion congruency effects. *Attention, Perception, & Psychophysics*, 78, 1087-1103.
- Hutchison, K. A., Moffitt, C. C., Hart, K., Hood, A. V. B., Watson, J. M., & Marchak, F. M. Measuring task set preparation versus mind wandering using pupillometry. (2020). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46, 280-295.
- Irons, J. L., Jeon, M., & Leber, A. B. (2017). Pre-stimulus pupil dilation and the preparatory control of attention. *PLoS ONE*, 12, e0188787.
- Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: From stimulus-response to script-report. *Trends in Cognitive Sciences*, 6, 333-9.
- Jack, A. I., & Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousness. *Cognition*, 79, 161-96.
- Jennings, J. R., van der Molen, M. W., & Steinhauer, S. R. (1998). Preparing the heart, eye, and brain: Foreperiod length effects in a nonaging paradigm. *Psychophysiology*, 35, 90-98.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices

- of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310-339.
- Kahneman, D. (1973). *Attention and effort*. Upper Saddle River: Prentice Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154, 1583-1585.
- Kane, M. J., Bleckley, K. M., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183.
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychological Science*, 18, 614-621.
- Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47-70.
- Kane, M. J., Smeekens, B. A., Meier, M. E., Welhaf, W. S., & Phillips, N. E. (2021). Testing the construct validity of competing measurement approaches to probed mind-wandering reports. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01557-x>
- Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley
- Keren, N. I., Taheri, S., Vazy, E. M., Morgan, P. S., Granholm, A-C, E., Aston-Jones, G. S., & Eckert, M. A. (2015). Histologic validation of locus coeruleus MRI contrast in post-mortem tissue. *NeuroImage*, 113, 215-245.
- Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health

- Organization adult ADHD Self-Report Scale (ASRS). *Psychological Medicine*, 35, 245-256.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- Lambie, J., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review*, 109, 219-259.
- Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The persistence of thought: Evidence for a role of working memory in the maintenance of task-unrelated thinking. (2012). *Psychological Science*, 23, 375-380.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience Reviews*, 17, 347-356.
- Madore, K. P., Khazenzon, A. M., Backes, C. W., Jiang, J., Uncapher, M. R., Norcia, A. N., & Wagner, A. D. (2020). Memory failure predicted by attention lapsing and media multitasking. *Nature*, 587, 87-91.
- 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, 196-204.
- Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *Neuroscience*, 11, 1-17.
- Meier, M., Smeekens, B., Silvia, P. J., Kwapi, T., & Kane, M. J. (2018). Working memory capacity and the antisaccade task: A microanalytic-macroanalytic investigation of individual differences in goal activation and maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 68-84.

- Moffitt, C. C. (2013). Working memory capacity and saccade performance across fixation delay: Attentional preparation or goal neglect? (Master's Thesis) Retrieved from <http://scholarworks.montana.edu/xmlui/handle/1/2685>
- Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Ohtsuka, C., Sasaki, M., Konno, K., Koide, M., Kato, K., Takahashi, J., Takahashi, S., Kudo, K., Yamashita, F., & Terayama, Y. (2013). Changes in substantia nigra and locus coeruleus in patients with early-stage Parkinson's disease using neuromelanin-sensitive MR imaging.
- Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Journal of Applied Psychology*, 59, 213-216.
- 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*, 88, 676-682.
- Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society for Neuroscience Abstracts*, 19, 974.
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, 7, 132289.
- Richer, F., & Beatty, J. (1987). Contrasting effects of response uncertainty on the task-evoked pupillary response and reaction time. *Psychophysiology*, 24, 258-262.

- Richer, F., Silverman, C., & Beatty, J. (1983). Response selection and initiation in speeded reactions: A pupillometric analysis. *Journal of Experimental Psychology: Human Perception & Performance*, 9, 360-370.
- Richmond, L. L., Burnett, L. K., Morrison, A. B., & Ball, H. (2021). Performance on the processing portion of complex working memory span tasks is related to working memory capacity estimates. *Behavior Research Methods*. DOI: 10.3758/s13428-021-01645-y.
- Robison, M. K., & Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception, & Psychophysics*, 82, 3273-3290.
- Robison, M. K., Miller, A. L., & Unsworth, N. (2019). Examining the effects of probe frequency, response options, and framing within the thought-probe method. *Behavior Research Methods*, 51, 398-408.
- Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics*, 81, 407-419.
- Rondeel, E. W., Van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in human neuroscience*, 9.
- Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, 21, 1309–1315.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, 8, 23-74.

- Schooler, J. W., (2002b). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences*, 6, 339-344.
- Schooler, J. W., & Schreiber, C. A. (2004). Experience, meta-consciousness, and the paradox of introspection. *Journal of Consciousness Studies*, 11, 17-39.
- Schubert, A L., Frischkorn, G. T., & Rummel, J. (2020). The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychological Research*, 84, 1846-1856.
- Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, 4, 1-6.
- Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process– occurrence framework for self-generated mental activity. *Psychological Bulletin*, 139, 519 -535.
- Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M., Carlson, J. M., & Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE*, 6, 1-8.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.
- Smallwood, J., Brown, K. S., Tipper, C., Gleasbrecht, B., Franklin, M. S., Mrazek, M. D., ... Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS One*, 6, e18298.
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, 132, 946-958.

- Sun, W., Tang, Y., Qiao, Y., Ge, X., Mather, M., Ringman, J. M., & Shi, Y. (2020). A probabilistic atlas of locus coeruleus pathways to transentorhinal cortex for connectome imaging in Alzheimer's disease. *NeuroImage*, *223*, 117301.
- Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49*, 110-128.
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635-654.
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the Antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 1302-1321.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavioral Research Methods*, *37*, 498-505.
- Unsworth, N., & Miller, A. L. (2021). Individual differences in the intensity and consistency of attention. *Current Directions in Psychological Science*. DOI: 10.1177/09637214211030266
- Unsworth, N., Miller, A. L., & Robison, M. (2020). Individual differences in lapses of sustained attention: Oculometric indicators of intrinsic alertness. *Journal of Experimental Psychology: Human Perception & Performance*, *46*, 569-592.
- Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review*, *22*, 757-765.
- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention.

Cognitive, Affective, and Behavioral Neuroscience, 16, 601-615.

- Unsworth, N., & Robison, M. K. (2017a). A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review*, 24, 1282-1311.
- Unsworth, N., & Robison, M. K. (2017b). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 1962-1987.
- Unsworth, N., Robison, M. K., & Miller, A. L. (2018). Pupillary correlates of fluctuations in sustained attention. *Journal of Cognitive Neuroscience*, 30, 1241-1253.
- Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, 22, 549-554.
- Van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS ONE*, 11, 1-16.
- 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(1), 158-169.
- van der Molen, M. W., Boomsma, D. I., Jennings, J. R., & Nieuwboer, R. T. (1989). Does the heart know what the eye sees? A cardiac/pupillometric analysis of motor preparation and response execution. *Psychophysiology*, 26, 70-80.
- Wang, C. A., Brien, D. C., & Munoz, D. P. (2015). Pupil size reveals preparatory processes in the generation of pro-saccade and anti-saccades. *European Journal of Neuroscience*, 41, 1102-1110.

Table 1.

Correlations between saccade accuracy, CEPRs, CEPR variability, AOSPAN, and TUTs.

	1	2	3	4	5	6	7	8	9	10	11
1. Antisaccade Accuracy	--										
2. Prosaccade Accuracy	.365	--									
3. AOSPAN	.323	.170	--								
4. Antisaccade early CEPR	-.269	-.163	-.228	--							
5. Prosaccade early CEPR	-.255	-.268	-.232	.791	--						
6. Antisaccade late CEPR	.455	.142	.087	-.116	-.156	--					
7. Prosaccade late CEPR	.355	.122	-.010	-.055	-.073	.807	--				
8. Antisaccade CEPR SD	-.310	-.308	.015	.190	.236	-.189	-.164	--			
9. Prosaccade CEPR SD	-.259	-.175	.045	.196	.230	-.148	-.099	.913	--		
10. Antisaccade TUT	-.261	-.039	-.016	.037	.052	-.300	-.318	.263	.240	--	
11. Prosaccade TUT	-.262	-.026	-.036	.010	.017	-.298	-.292	.250	.263	.840	--

Note. Significant effects ($p < .05$; two-tailed) are in bold.

Table 2.
Results of Regression Analyses Predicting Saccade Accuracy.

Predictor	<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>t</i>	<i>F</i>	<i>Adj. R2</i>
Antisaccade Accuracy					15.609	.363
AOSPAN	.002	.001	.257*	3.531		
Early CEPR	-.219	.103	-.158*	2.122		
Late CEPR	.550	.117	.356*	4.696		
Pupil SD	-.156	.070	-.173*	2.222		
TUT rate	-.047	.033	-.109	1.406		
Prosaccade Accuracy					3.626	.093
AOSPAN	.001	.001	.138	1.584		
Early CEPR	-.214	.100	-.192*	2.138		
Late CEPR	.172	.148	.103	1.164		
Pupil SD	-.131	.064	-.191*	2.041		
TUT rate	.031	.031	.092	0.980		
Overall Accuracy					12.582	.311
AOSPAN	.001	.000	.259*	3.397		
Early CEPR	-.188	.087	-.170*	2.159		
Late CEPR	.377	.110	.270*	3.418		
Pupil SD	-.182	.056	-.269*	3.272		
TUT rate	-.004	.027	-.012	0.143		

Note. * = $p < .05$; two-tailed

Figure 1.

Sample Antisaccade Trial Sequence for Normal and Thought Probe Trials.

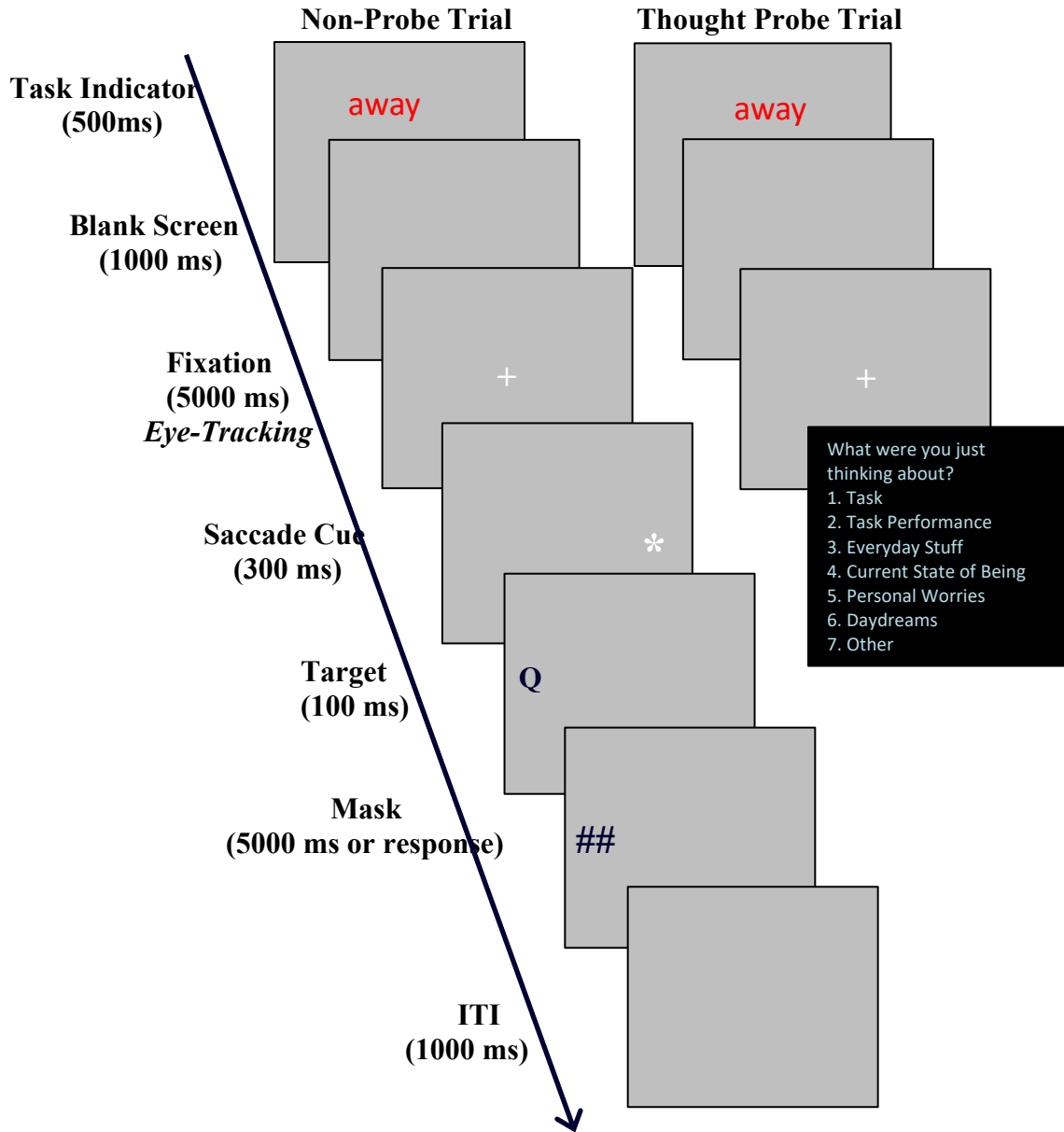


Figure 2.

Change in pupil diameter during 5000 ms fixation period as a function of cued trial type. Error bars reflect standard error for paired-sample difference across time periods.

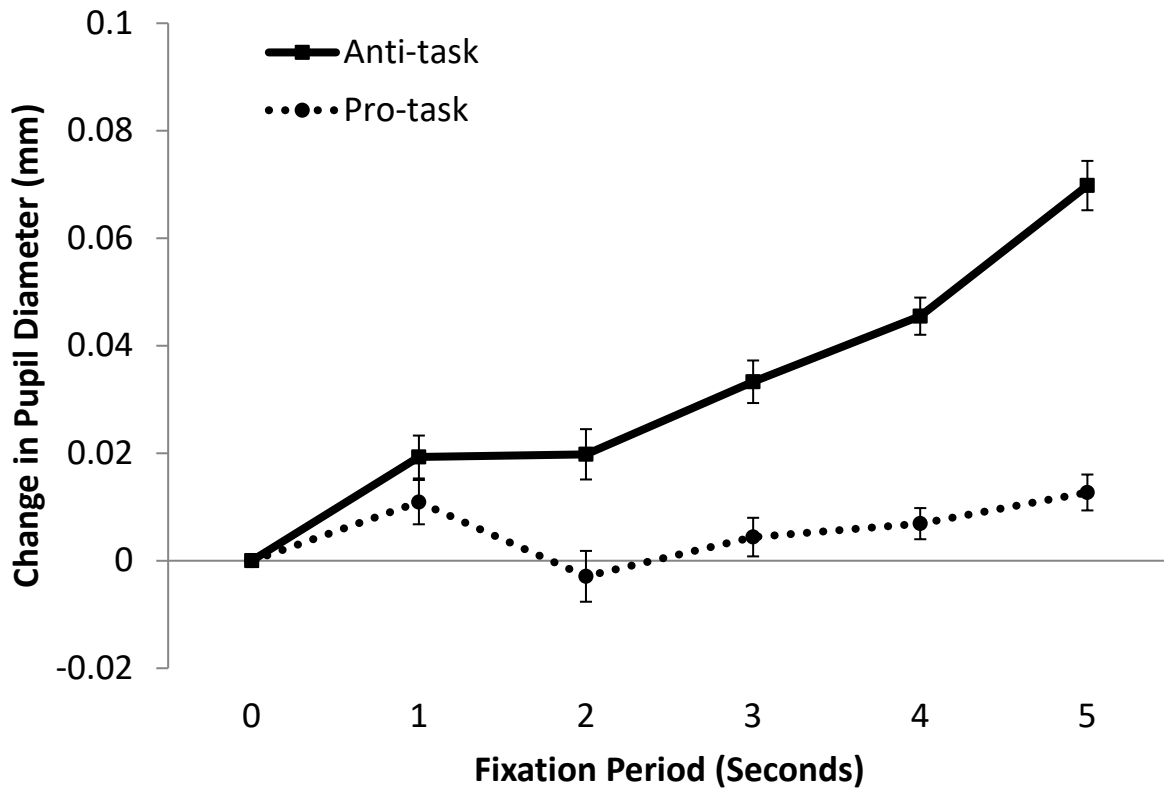


Figure 3.

Change in pupil diameter during 5000 ms fixation period for antisaccade trials preceding accurate (solid) and error (dotted) responses. Error bars reflect standard error for paired-sample difference across time periods.

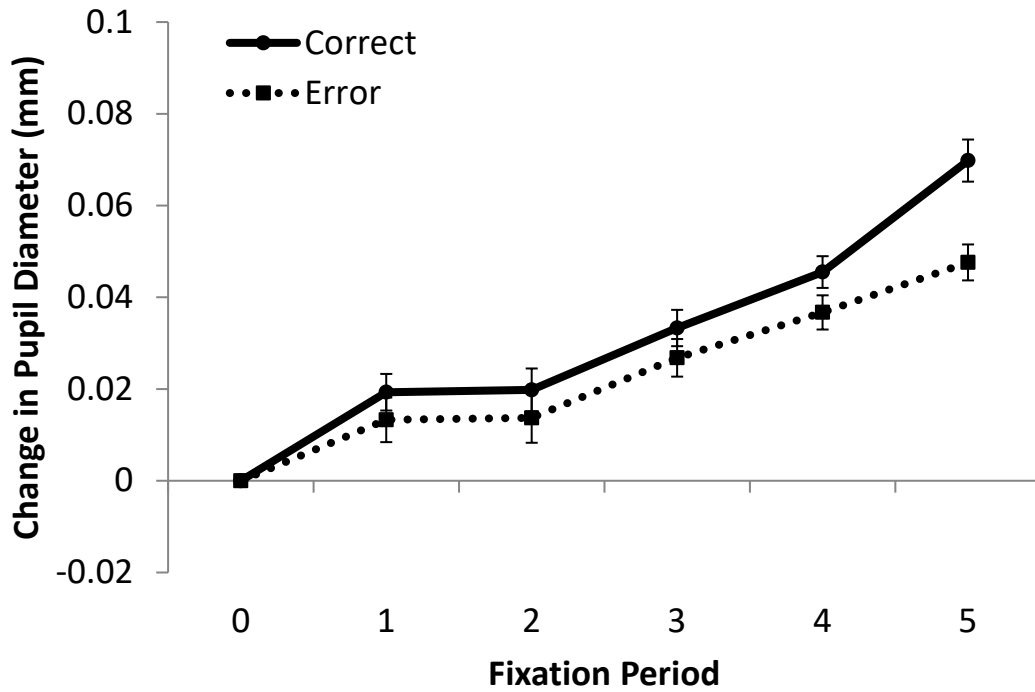


Figure 4.

Change in pupil diameter during 5000 ms fixation period for antisaccade trials (top) and prosaccade trials (bottom) from Good (solid), Middle (dashed), and Poor (dotted) performers in each task. Error bars reflect standard error for paired-sample difference across time periods.

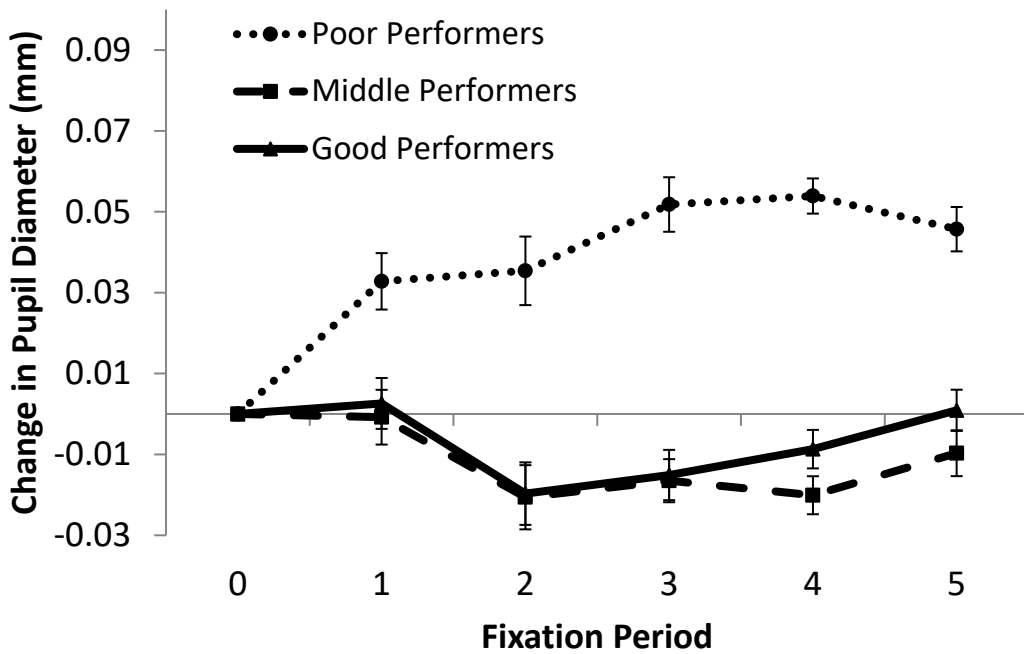
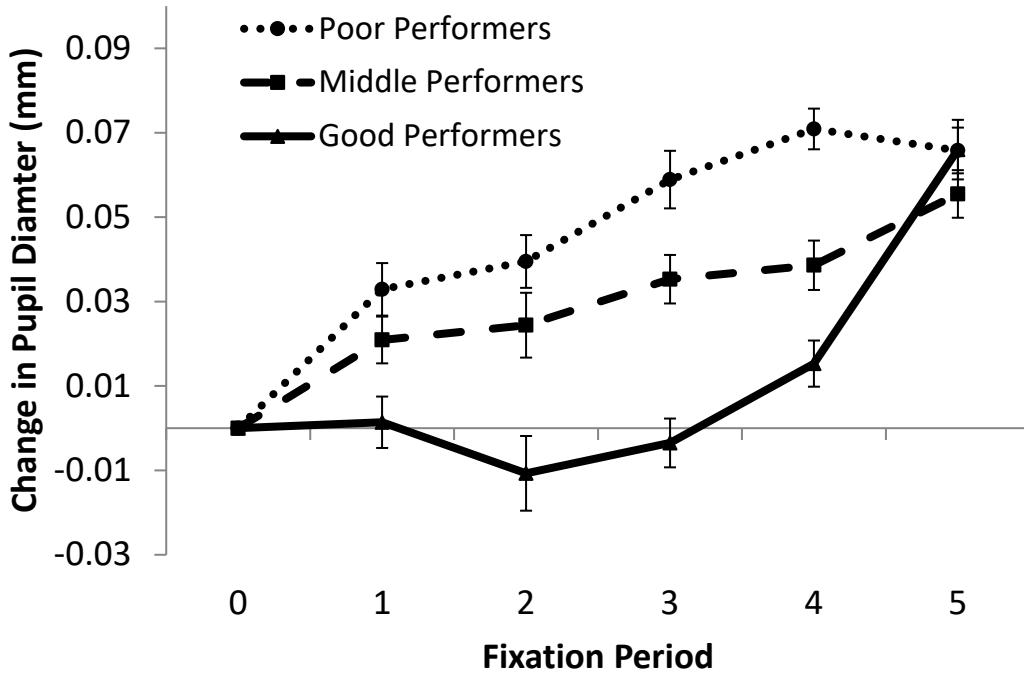


Figure 5.

Change in pupil diameter during 5000 ms fixation period for antisaccade trials (top) and prosaccade trials (bottom) from High WMC (solid), Middle WMC (dashed), and Low WMC (dotted) participants. Error bars reflect standard error for paired-sample difference across time periods.

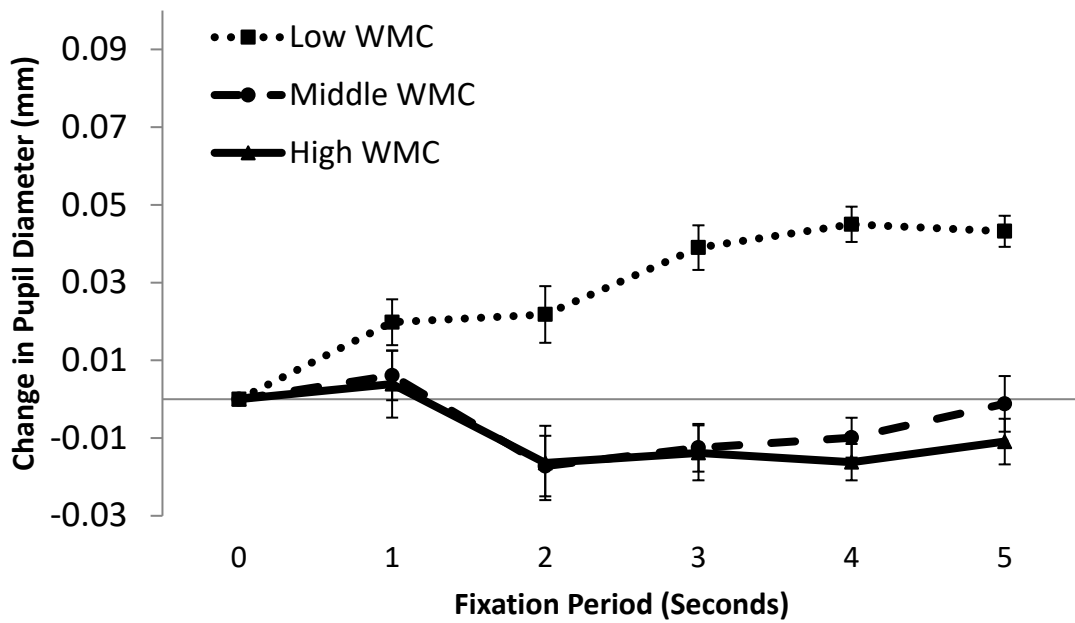
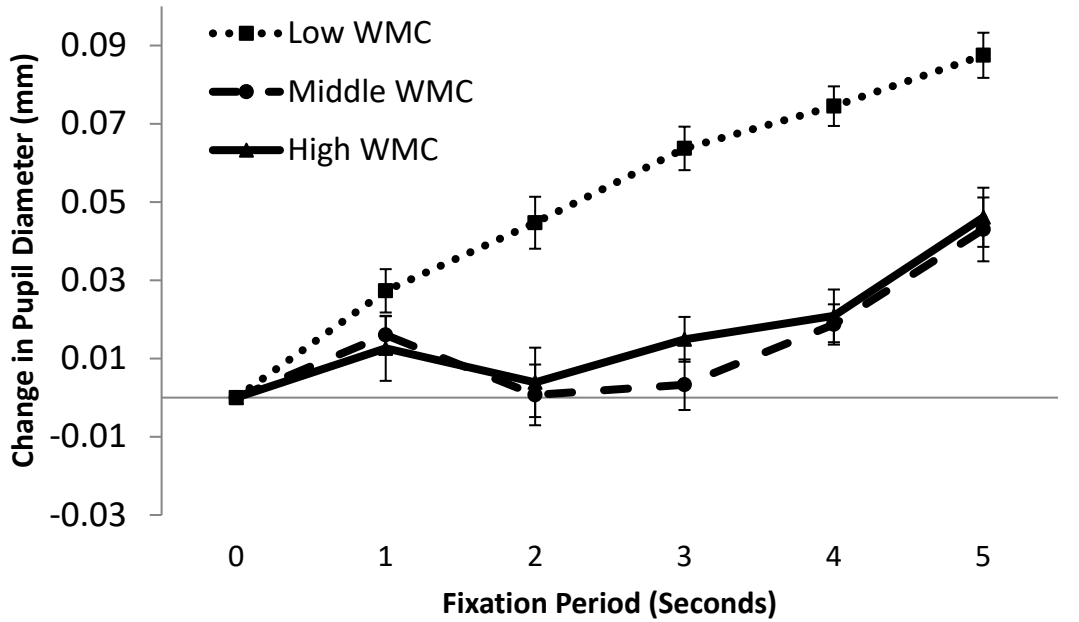


Figure 6.

Mean self-reported TUT rate on prosaccade and antisaccade trials from Hutchison et al. (2020, Experiments 1 & 2) and the current Experiment. Error bars reflect standard error for paired-sample difference between trial types within each experiment.

