PUPILLARY RESPONSES TO SIMPLE DIFFICULTY CUES

A Thesis
Presented to
The Academic Faculty

by

Alexandra Trani

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Psychology

Georgia Institute of Technology

December 2016

COPYRIGHT © 2016 BY ALEXANDRA TRANI

PUPILLARY RESPONSES TO SIMPLE DIFFICULTY CUES

Approved by:

Paul Verhaeghen, Advisor
School of Psychology
Georgia Institute of Technology

Audrey Duarte
School of Psychology
Georgia Institute of Technology

Keith Hutchison
School of Psychology
Montana State University

Date Approved: April 7, 2016

ACKNOWLEDGEMENTS

Thank you to Drs. Paul Verhaeghen, Audrey Duarte, Keith Hutchison, Mark Scott, and Jason Watson – your timely encouragement and sagely advice helped make this thesis a reality. A very warm thank you to Adrian Braicu, Christine Trani, Jon Strunk, and Christine Godwin for the hours you spent listening to me rant about data.

LIST OF TABLES

Table 1 -- Correlations between cognitive ability variables and accuracy. 21

Table 2 -- Correlations between cognitive ability variables and accuracy. 24

Table 3 -- Univariate results for the Garavan linger analyses. 25

LIST OF SYMBOLS AND ABBREVIATIONS

PDR Pupil dilation ratio

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES iv

LIST OF SYMBOLS AND ABBREVIATIONS v

SUMMARY vii

CHAPTER 1. Introduction 1
1.1 Phasic pupil size 3
1.2 Emotional Arousal and the Pupillary Response 3
1.3 Pupillary responses to difficulty 4

CHAPTER 2. Hypotheses and Potential Interpretations 8
2.1 Difficulty Cues 8
2.2 End of Trial Fixations 10
2.3 Summary 11

CHAPTER 3. Methods 13
3.1 Participants 13
3.2.1 Dark/Light Adaptation Task. 14
3.2.2 Eyetracker Task Commonalities. 14
3.2.3 Sound-Based Garavan Task. 15
3.2.4 Visual Search Task. 16
3.2.5 Cognitive Ability and Need for Cognition. 17
3.3 Pupillary Data Filtering 18

CHAPTER 4. Results 20
4.1 Correlations 20
4.2 Garavan Full Analyses 21
4.3 Garavan Accuracy Analyses 23
4.4. Garavan Linger Analyses 24
4.5 Visual Search Full Analyses 25
4.6 Visual Search Complete Accuracy Analyses 26
4.7 Visual Search Linger Analyses 29

CHAPTER 5. Discussion 30
5.1 Findings for the Sound-Based Garavan 31
5.2 Findings for the Visual Search Task 33
5.3 Could People Infer Task Performance In The Visual Search Task? 34
5.5 Limitations 36

REFERENCES 38

SUMMARY

Increases in emotional arousal and attention control simultaneously enhance activity in the sympathetic nervous system and cause pupils to dilate. Additionally, the pupil has been shown to be sensitive to changes in task demands – changes that would also influence emotional arousal and attention control. Difficulty cues may improve performance by helping people efficiently prepare to meet task demands. In this study, I presented people with simple difficulty cues (i.e., easy, medium, or hard) to elicit preparatory attention control. I hypothesized that pupil size would be larger while viewing cues indicative of harder trials *and* this difference in pupil size would be associated with better performance. Pupil size during the difficulty cue was unrelated to accuracy and appears to be the same despite differences in cue content. I hypothesized that end-of-trial pupil size would increase with trial difficulty *and* differences in pupil size may relate to trial performance. End-of-trial pupil size increased with difficulty and was found to relate to trial performance.

CHAPTER 1. INTRODUCTION

The ability to prioritize incoming information, ignore distractions, and focus cognition on task or goal-relevant demands is referred to as attention control (Lange & Schnuerch, 2014). While behavioral measures, such as reaction time and accuracy, offer some insight into a person's underlying ability to control attention and meet task demands, incorporating physiological measures provides a more complete understanding of the relationship between brain function, task performance, and time.

There are three main physiological measures frequently used to study the effects of cognitive load. First, functional magnetic resonance imaging has been used to show how changes in activation across brain areas work to alter performance (Ester, Vogel, & Awh, 2012; Franco, 2010). Second, electroencephalography and event-related potential techniques measure summations of post synaptic potentials, revealing precision at the millisecond level of when cognitive events have occurred (Slagter, Prinssen, Reteig, & Mazaheri, 2016). Finally, pupillometry, the study of small scale changes in pupil size functionally unrelated to vision, provides an online index of attentional engagement and arousal (Beatty & Lucero-Wagoner, 2000).

The present study used pupillometry to determine the extent to which a person can ready attention in preparation for an upcoming trial in response to the presentation of a difficulty cue. Should people be able to utilize the cues, I hypothesized that harder cues would elicit larger pupil sizes than easier cues because people would focus

attention in preparation for the start of the trial to better meet task demands and maintain or increase performance.

Pupillary dilations as measured by pupillometry reflect activation of the sympathetic nervous system through the release of norepinephrine in the forebrain by fibers tracing back exclusively to the locus coeruleus (Aston-Jones & Cohen, 2005; Gilzenrat, Cohen, Rajkowski, & Aston-Jones, 2003; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Laeng, Sirois, & Gredeback, 2012; Phillips, Szabadi, & Bradshaw, 2000; Rajkowski, Kubiak, & Aston-Jones, 1993). Activation in the locus coeruleus increases sympathetic activity and reduces parasympathetic activity by activating $alpha_1$- and $alpha_2$-adrenoceptors respectively (Lewis & Coote, 1990; Smith, Schambra, et al., 1994). As activity in the locus coeruleus and sympathetic nervous system increases, so too does a person's overall arousal level (Samuels & Szabadi, 2008) and, because pupils can dilate with latencies of 200 milliseconds (Lowenstein & Loewenfeld, 1962), all of this is reflected in pupillary dilations (Lowenfold, 1993; Steinhauer et al., 2004). Finally, pupil size in humans correlates with performance on tasks linked to norepinephrine release by the locus coeruleus (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Jepma & Nieuwenhuis, 2011; Murphy, Robertson, Balsters, & O'Connell, 2011).

There are two types of pupillary dilations commonly studied in the literature: tonic and phasic dilations. *Tonic pupil size*, not measured in the present study, is a person's average pupil size throughout the day and is usually experimentally measured with a pre-experimental baseline. *Phasic pupillary dilations* are discussed in detail in the following section.

1.1 Phasic pupil size

Phasic pupil sizes, or task-evoked pupillary responses, are a person's average pupil size in response to a particular event such as a trial in an experiment or, in the case of the present study, the anticipation of an upcoming task of a particular difficulty level. Task-evoked phasic pupil dilations provide an online index of fluctuations in attention within a task at a given moment in time (Just & Carpenter, 1993). Because of this, phasic pupil size is assumed to reflect environmental exploitation, that is, selective attention to and processing of goal-relevant stimuli that are interesting and/or rewarding in the context of a task (Aston-Jones & Cohen, 2005). Phasic pupillary dilations occur regardless of actually receiving an award and are separate from dilations associated with response preparation (Aston-Jones, Rajkowski, Kubiak, & Alexinsky, 1994). The present study examined the relationship between changes in phasic pupil size, accuracy, and the presence of a difficulty cue.

1.2 Emotional Arousal and the Pupillary Response

Pupil dilation is indicative of cognitive effort and emotional responses to arousing material. The seminal study here is by Hess and Polt (1960), who examined how pictures varying in emotional valence affected the pupillary response. While no statistical test was performed to confirm the significance of these findings, pupils constricted while people viewed unpleasant pictures and dilated while people viewed pleasant pictures. Because this study used only five pictures and had a small sample size, initial attempts to replicate were confusing if not outright conflicting (Libby, Lacey, & Lacey, 1973).

More recent research suggests that the pupil dilates to emotional stimuli regardless of valence. Specifically, pupil diameter is larger for both negatively and positively valenced photographs when compared to neutral photographs (Aboyoun & Dabbs, 1998; Steinhauer, Boller, Zubin, & Pearlman, 1983). These findings extend beyond photographs; measured pupil diameter is larger for emotionally valenced words than for neutral words (Bradley, Miccoli, Escrig, & Lang, 2008; Janisse, 1974, Võ, Jacobs, Kuchinke, Hofmann, Conrad, Schact, & Hutzler, 2008). Even emotionally valenced sounds evoke larger pupil sizes than neutral sounds (Partala, Jokiniemi, Surakka, 2000; Partala & Surakka, 2003).

1.3 Pupillary responses to difficulty

Studies testing the relationship between task demands and pupil size began as early as 1875 (Schiff, 1875). By 1896, Heinrich reported larger pupil diameters for people solving more difficult multiplication problems than those people solving easier problems; a finding that has since been replicated using more advanced equipment (Hess & Polt, 1964). Specifically, in more difficult conditions demanding more controlled attention, the pupil dilates faster and to a larger size while maintaining peak dilation for longer times than in easier conditions (for replications, see Hess & Polt, 1964; Jiang, Zheng, Bednarik, & Atkins, 2015; Kahneman & Beatty, 1966a, 1966b, 1967; Payne, Parry, & Harasymiw, 1968; Porter, Troscianko, & Gilchrist, 2007). The pupil follows a specific time course when solving multiplication problems; reaching peak dilation at the point of solution and returning to baseline upon report of the answer (Bradshaw, 1968; Hess & Polt, 1964; Kahneman & Beatty, 1966b). These findings

suggest a link between attention control and pupil size as more attention would be needed to solve more difficult trials.

This effect is not limited to multiplication tasks as the pupil reliably dilates as a function of real and perceived task difficulty across domains (Jiang, Zheng, Bednarik, Atkins, 2015; Nuthmann & van der Meer, 2005; Raisig, Welke, Hagendorf, & van der Meer, 2007; Verney, Granholm, & Marshall, 2004). In a visual search task, pupil dilations were sensitive enough to be the only measure that distinguished difficulty conditions (Porter, Troscianko, & Gilchrist, 2007). Kuchinsky and colleagues (2013) presented older adults with a word identification task consisting of (a) two levels of signal to noise ratio: easy and hard, (b) and a distractor word that varied in phonological similarity to the target word and was played simultaneously with the target word. Pupil size was found to be largest for the most challenging conditions.

Moresi and colleagues (2008a) used a finger-cuing task to present people with cues that differed in motor response preparation difficulty (i.e., uncued, hand-cued, finger-cued, cuing neither finger nor hand) in what may be the only two pre-existing studies examining pupillary responses to difficulty cues. Larger pupillary dilations occurred as motor-response preparation difficulty increased (Moresi et al., 2008a), nevertheless, these results may be specific to motor preparation. This thesis tested if difficulty cues not associated with a particular response type could evoke changes in emotional arousal and attentional control that may help (or hinder) task performance.

As tasks become automated and less controlled attention is needed to maintain performance, pupillary responses decrease strongly suggesting that the pupil is reflecting controlled attention (Kahneman & Beatty, 1967). Finally, phasic pupillary

responses also reveal limits in attention control. Increased attention results in larger pupil sizes until a person has reached the limit of their ability to attend, at which point the pupil may begin to constrict (Peavler, 1974; Granholm, Asarnow, Sarkin, & Dykes, 1996). It could be assumed that if these results were due to emotional arousal alone, pupil size would not decrease at the hardest levels; instead, the hardest difficult might evoke larger pupillary sizes as people become more emotionally aroused (e.g., frustration or anxiety) by the task.

There are several pieces of evidence pointing toward a relationship between the phasic pupillary response and strategic attention control as it relates to memory encoding. First, larger pupil sizes are observed for items more difficult to maintain in or retrieve from memory – that is, items that require greater attention control (e.g., a longer digit string or list compared to a shorter string or list; van Rijn, Dalenberg, Borst, & Sprenger, 2012). With every item presented in a short term memory task, there is a corresponding increase in pupil size because more attention is necessary to hold each additional item in memory (loading phase; Kahneman & Beatty, 1966; Peavler, 1974; Granholm, Asarnow, Sarkin, & Dykes, 1996). The pupil constricts with the report of each remembered item because less attention is needed to maintain the string in memory (unloading phase; Kahneman & Beatty, 1966; Peavler, 1974; Granholm, Asarnow, Sarkin, & Dykes, 1996). Once the final item is reported and attention begins to wander, the pupil will return to baseline (Kahneman & Beatty, 1966; Peavler, 1974; Granholm, Asarnow, Sarkin, & Dykes, 1996). In a paired-associates task, pupil diameter was consistently smaller during retrieval for those pairs that were better

learned and thus had stronger memory traces (van Rijn, Dalenberg, Borst, & Sprenger, 2012).

Assuming that the pupil dilates to increases in both emotional arousal and attention control, is the pupil still sensitive to increases in attention control in the face of potential emotional arousal? Stanners, Coulter, Sweet and Murphy (1979) found that if task demands required a considerable amount of mental effort, then pupillary responses reflected cognitive processing rather than emotional arousal. These findings have been recently replicated using a lexical decision task where it was the frequency, not the emotional content, of the words that evoked pupillary responses (Kuchinke, Võ, Hofmann, & Jacobs, 2007).

When considering the above findings, it is easy to understand why Bumke (1911) argued that the pupillary system was so sensitive to changes in cognition that every mental event could be observed through fluctuations in pupil size. While the pupil may not reveal every mental event, increases in attention control, task difficulty, and emotional arousal all cause reliable changes in pupil size; changes that can be used to make relatively stable predictions. Using what we know about attention control, emotional arousal, and the pupillary response, I made the following predictions with my thesis.

CHAPTER 2. HYPOTHESES AND POTENTIAL INTERPRETATIONS

The present study measured pupil size during a difficulty cue (i.e., easy, medium, or hard) presented at the beginning of each trial and pupil size at the end of each trial during a fixation for (1) a working memory task (i.e., a sound-based Garavan task) and (2) a visual search task. These two tasks allowed me to compare two potentially different types of strategic control; visual search tasks may evoke preparation of more perceptual strategies while the sound based Garavan task may evoke preparation of strategies more typical of what is expected in an attention control task. Pupil size during the difficulty cue likely reflects emotional engagement or attention evoked by the presence of the difficulty cue *before the trial begins*. Pupil size during the end of trial fixation likely reflects changes in emotional arousal or attention that *resulted from the trial itself*. The next two sections will detail hypothesized results as well as their interpretations.

2.1 Difficulty Cues

I will examine two main effects for pupil size during the difficulty cue. The first main effect, cue, examines whether pupil size during the difficulty cue was significantly different across easy, medium, or hard cues. The second main effect, subsequent accuracy, examines whether pupil size during the difficulty cue was significantly different for trials answered correctly when compared to trials answered incorrectly. For the main effect of cue, pupil size may be larger for difficulty cues preceding harder trials than for difficulty cues preceding easier trials. I believe this to be the most likely hypothesis since it suggests that people are preparing for the trial using a potential

combination of increased attention control and emotional arousal. Since hard trials are very difficult, I believe that people will experience increased emotional arousal at the very least for cues preceding hard trials when compared with other difficulty cues. In contrast, pupil size may be smaller for difficulty cues preceding harder trials than for difficulty cues preceding easier trials suggesting that people are perhaps emotionally disengaging from more difficult trials. Lastly, pupil size may not be significantly different across difficulty cues regardless of cue content suggesting that people were not using the cue to prepare for trials.

For the main effect of accuracy, pupil size during the difficulty cue for correct trials may be significantly larger or smaller than for incorrect trials. If pupil size during the difficulty cue for correct trials is significantly larger than incorrect trials, this would suggest that people are either more emotionally engaged or are preparing attention in such way as to meaningfully impact performance. If pupil size during the difficulty cue for incorrect trials is significantly smaller for correct trials, then people may be preparing attention in a more streamlined or efficient way that leads to better performance. Additionally, smaller pupil size during difficulty cues for correct trials may suggest that people are less emotionally aroused (e.g., less anxious) and this may lead to better performance. Lastly, there is the possibility that pupil size during the difficulty cue for correct trials is not significantly different from incorrect trials which would suggest that people were not using the cue in a way that would meaningfully affect performance.

While main effects are interesting, it is the presence of an interaction between cue and subsequent accuracy that would uniquely point to an effect caused by

preparatory attention control. If people were able to use the difficulty cue to prepare attention in a meaningful way so as to impact trial accuracy, I would expect the interaction between cue and subsequent accuracy to be significant. If neither the interaction nor the main effect of subsequent accuracy is significant, this suggests people were unable to prepare attention sufficiently to cause changes in trial accuracy.

2.2 End of Trial Fixations

Pupil size during the end of trial fixations may reveal important information about the time it takes for the pupil to return to a baseline level. Patterns in pupil size here may reveal effort related issues such as voluntarily giving up during more difficult trials.

In regards to the main effect of accuracy, there are three distinct hypotheses that would hold despite difficulty level. If pupil size is larger at the end of correct trials than at the end of incorrect trials, this could suggest that more attention control was necessary to answer more difficult trials. This pattern of interaction could also suggest that people are somehow aware of their performance and are more emotionally aroused (e.g., surprised) when they answer more difficult trials correctly – this is especially plausible for the visual search task because people were inadvertently aware of trial performance. If pupil size is smaller at the end of correct trials than at the end of incorrect trials, this could suggest that people were more emotionally aroused (e.g., frustrated) at the end of trials answered incorrectly – a scenario that is especially plausible for the visual search task where performance feedback was inadvertently provided.

In regards to the main effect of difficulty, I would expect pupil size to increase with difficulty except in cases of cognitive overload. Cognitive overload may cause pupil size to decrease at the end of more difficult trials because people are giving up or have reached the limits of their mental ability. If pupil size is larger at the end of more difficult trials, this could suggest that people used more attention control and/or experienced higher emotional arousal during more difficult trials that lingered to the end of the present trial. If pupil size is smaller at the end of more difficult trials, this could suggest that people weren't trying as hard for more difficult trials – that is, people were less emotionally engaged or were less attentive during difficult trials – a pattern that is especially probable if difficulty exceeds ability.

Finally, there could be a significant interaction between pupil size, difficulty, and accuracy. If pupil size is larger for more difficult trials *and* for accurate trials than for easier inaccurate trials, this could suggest that people utilized more attention control to solve more difficult trials correctly. There is another possibility: if people were somehow aware of their performance, it could be that people were more emotionally aroused at the end of more difficult, correct trials than at the end of easier trials or those trials answered incorrectly – this is especially plausible for the visual search task. If pupil size is larger for more difficult trials *and* for incorrect trials than for easier or correct trials, this could suggest that people were more emotionally aroused or increased attention control was not sufficient to influence performance.

2.3 Summary

Considering that pupillary responses only reflect emotional arousal when cognitive task demands are minimal (Stanners, Coulter, Sweet, Murphy, 1979), end-of-

trial pupil size could reflect differences in both emotional arousal and attention control because the end-of-trial task was not cognitively demanding (staring at a fixation cross). Pupil size during the difficulty cue is more complicated; it is possible that viewing the cue is a cognitively demanding task if people used this time to prepare attention or rehearse strategies. Only if the interaction between cue and subsequent accuracy proved significant would I say with confidence that people engaged in preparatory attention control processes. While it is certainly possible that people engaged in preparatory attention processes that were simply not sufficient to improve subsequent accuracy, I cannot speak to the extent that pupillary responses were caused by attention control rather than emotion if the interaction between cue and subsequent accuracy proves to be non-significant.

CHAPTER 3. METHODS

3.1 Participants

I recruited 110 students attending the Georgia Institute of Technology. Participants were screened for corrected to normal hearing and vision. Participants were instructed to refrain from consuming substances like nicotine, caffeine, and alcohol for at least four hours before the study to help ensure that pupil size was not influenced by substance use. Participants who indicated drinking more than a standard cup of coffee less than two hours before their scheduled study session were removed from the present analyses. Additionally, participants over the age of 24 were filtered from my original dataset. After removing those students who did not meet these requirements from my dataset, I had a total of 96 students aged 18 to 24 (M = 19.47, SD = 1.37; 45 females).

All participants were tested in the same, light controlled room. SR Research's EyeLink II (500Hz) was used to record pupil-size data in arbitrary units based on the number of camera pixels occluded by the pupil image as well as a gamut of other measures including gaze position, blinks, and the number and duration of fixations – all determined by the EyeLink host computer software (SR Research, Ontario, Canada). Error variance in measurement can be due, among other things, to camera angle, distance from the eye to the camera, and individual differences in eye shape and size – all of these variables making the arbitrary units difficult to convert into an absolute unit of size (SR Research, Ontario, Canada). Considering this variance, a person's pupil size will generally range from 800-2000 units (Heaver, 2012).

3.2 Design

3.2.1 Dark/Light Adaptation Task.

Participants began the study with a five-minute dark adaptation session where they were instructed to stare at a dark purple fixation cross in the center of a black screen. Seven seconds before completion, the screen changed to a brilliant white screen with a light gray fixation cross in the center. Participants were warned about this change.

3.2.2 Eyetracker Task Commonalities.

Following the dark/light adaptation task, the eyetracker tasks consisted of (1) a sound-based Garavan task and (2) a visual search task. See descriptions below. The Garavan task is a cognitively difficult task tapping working memory while the visual search is a perceptually difficult task. While these tasks are very different in what they measure, they share commonalities. All trials were self-started. A self-started design allows participants to begin the next trial when ready and helps to prevent mind wandering. Additionally, the eyetracker tasks had items with three levels of difficulty: easy, medium, or hard.

At the beginning of each task and before reading task instructions, I presented a seven second task baseline measure that consisted of a single fixation cross in the center of a black screen. Participants were told to stare at the fixation cross for the entire presentation time. Pupil dilation in this baseline was used as the denominator in the calculation of percent change in dilation for each trial.

A difficulty cue was presented for three seconds before each trial began to alert people of the present trial's difficulty level. Difficulty cues consisted of a single, uppercase word (easy, medium, or hard) appearing in the center of the screen

immediately after the subject starts the trial. Cues in the present study were 100% valid. Before each task, participants were informed of the meaning of the difficulty levels. Easy trials were described as trials where "most people are able to complete the trial with little challenge". Medium trials were described as trials "some people find to be very simple while other people may struggle to complete". Hard trials were described as trials that "the majority of people find to be quite challenging." Participants were instructed to "do the best they can... Answer as quickly and accurately as possible… regardless of the difficulty level". Pupillary responses were recorded for the 3000 milliseconds duration of the difficulty cue. The trial began following the difficulty cue.

At the end of each trial, I presented a two-second black screen containing a single fixation cross in the center. Participants were instructed to stare at the fixation cross for the entire time of presentation. This end-of-trial fixation was presented to allow pupil size to return to baseline. Additionally, I hypothesized pupillary responses at the end of trial fixations might differ between correct and incorrect trials in a way similar to error related negativity or post error positivity seen in electroencephalogram and event-related potential experiments.

3.2.3 Sound-Based Garavan Task.

The Garavan task (1998) is believed to measure memory updating. For the present experiment, I used a sound-based Garavan task. Participants were presented with a sequence of *N* different sounds (with *N* ranging from 2 to 4). I used four, one-second sounds in the present study: a phone ring back tone, a low-pitched bell tone, a bird singing, and a wolf howl. Background noise and auditory pops were edited out of the sounds. Participants were provided with a single pair of Sennheiser HD201 over-

ear headphones to complete the task. Participants were asked to keep a running count of how many sounds of each type were presented.

In the easy condition, there were two distinct sounds to keep track of. In the medium condition, there were three distinct sounds to keep track of. In the difficult condition, there were four distinct sounds to keep track of. Series of sounds varied in length from 11 to 15 items. In total, there were 60 items (20 from each level of difficulty; four trials in each list length of a particular difficulty). These items were pseudorandomized so that roughly equal numbers of each condition appeared in 2 blocks of thirty items (i.e., for every 30 items, 1/3 of the trials were easy, 1/3 of trials were average, and 1/3 of trials were difficult). There was a short interlude between blocks where I re-calibrated the eye-tracker.

After the difficulty cue was presented, participants were told which sound types would play in the upcoming trial (e.g., phone and wolf). In order to ensure a steady pupillary response, sounds were played for 1000 milliseconds and followed by 500 milliseconds of silence before the next sound began (i.e., sounds advanced automatically). At the end of each sound series, participants used the mouse to enter in the number of sounds of each type they heard. Accuracy feedback, provided during practice, was not provided during the actual task to prevent participants from altering behavior and/or getting frustrated with the task.

3.2.4 Visual Search Task.

In the visual search task, participants were instructed to find the “P” within a 40 x 40 array of distracter letters. In easy conditions, “I” was the distracter letter. Letters “T” and “R” were distractors in the medium conditions. Letters “B” and “R” were

distracters in the hard conditions. Each array was available until either a person selected the target letter "P" or 10 seconds had passed. While accuracy feedback was not directly provided, participants knew that each array contained the target letter “P”. Because of this, participants had some knowledge about their performance – participants who could locate the target knew they answered correctly while participants who could not locate the target knew they answered the trial incorrectly.

The visual search task consisted of 300 items, 100 for each level of difficulty (easy, medium, and hard). These items were pseudorandomized so that roughly equal numbers of each condition appeared in blocks of hundred items (i.e., for every 100 items, 1/3 of the trials were easy, 1/3 of trials were medium, and 1/3 of trials were difficult).

Originally, the visual search task was a single task with calibration screens every 100 trials to ensure that the cameras were still aligned. Due to excessive crashing, the visual search task was split into ten blocks of thirty trials (all presented in the same order as the original visual search). Because I did not provide any break between blocks, there is no reason to believe that participants in the 10-block condition would have performed any differently than participants in the full condition.

3.2.5 Cognitive Ability and Need for Cognition.

To better understand the relationship between pupil size and cognitive ability, I included the odd problems of Raven’s Progressive Matrices (thought to measure fluid intelligence) and the automated Operation Span Task (thought to measure working memory capacity; Unsworth, Heitz, Schrock, & Engle, 2005). Participants also completed the 18-item Need-For-Cognition task (Cacioppo, Petty, Feinstein, Jarvis,

1996; Cacioppo, Petty, Kao, 1984) allowing me to test if high performers were simply more mentally engaged in the task. I had originally hypothesized that there would be a correlation between need for cognition and cognitive ability, however, a significant correlation did not exist between these variables. For this reason, I used a composite z-score of the two cognitive ability measures (automated Operation Span Task and the odd problems on Raven's Progressive Matrices) to form a single cognitive ability factor and compared Need-For-Cognition task performance separately in order to gain a better understanding of how cognitive ability and mental engagement may alter pupil size.

3.3 Pupillary Data Filtering

Recall that the eyetracker returned pupillary size in arbitrary units. To make the arbitrary units a meaningful measurement, Maw and Pomplun (2004) formulated the pupil dilation ratio (PDR) obtained by dividing maximum trial pupil size by a single baseline measurement taken at the start of an experiment. To account for the fact that participants engaged in eye-tracker tasks for at least two hours and, over time, autonomic arousal naturally decreases causing decreases in baseline pupil size (Woodmansee, 1966) as well as iris muscle fatigue (Lowenstein & Loewenfeld, 1964), a task baseline (see: Eyetracker Task Commonalities) was used instead of an experimental baseline to calculate the PDR. Thus, the present thesis calculated PDR by dividing the maximum pupil size in a given trial by the mean pupil size during the task baseline (Heaver, 2012).

Blinks were filtered from the analysis using SR Research's Data Viewer software. Additionally, 103 milliseconds before the blink as well as 162 milliseconds following the blink were excluded to account for the occlusion of the pupil caused by

the eyelid during the process of the blink (Tucker & Johns, 2005). Because I used the maximum pupil size for difficulty cues and end of trial fixation calculations (as opposed to the mean pupil size during these times), removing blinks was only important for the task baseline pupil size and should not have influenced a person's maximum pupil size.

Considering the length of the experiment as well as individual differences in arousal, it was important to control for the amount of time a person had their eyes shut during a given fixation point. I used the number of blinks as well as the average blink duration to filter out sleep trials or trials with excessive blinking. I considered a person "asleep" during a given fixation point if the number of blinks multiplied by the average blink duration was at least 50% of the duration of the fixation. For example, I considered a person "asleep" during a 3000 millisecond difficulty cue if their eyes were shut for a total of 1500 milliseconds or more during a trial. If a person was found to be sleeping during 1/3 of the total fixation points for a given condition (i.e., during cue or end of trial fixation) in a task, the person was removed from further analyses. Eight participants were removed for this reason.

Due to data loss from crashes in the visual search, there are a different amount of subjects in the Garavan task analyses as opposed to the visual search task analyses. Garavan task analyses consisted of 95 possible subjects. Visual search task analyses consisted of 89 possible subjects.

CHAPTER 4. RESULTS

I have an entirely within-subjects (i.e., no between subjects factor), doubly-multivariate, repeated-measures design. Rather than running multiple ANOVAs, MANOVAs were run because this method offers greater power, remains robust against violations of sphericity assumptions, and helps to prevent type I error when analyzing the multiple dependent variables. My design theoretically violates Mauchley's sphericity assumption which states that all the different levels of the dependent variables should have equal variances between them. Because easy conditions subtracted from medium conditions cannot be assumed to have the same variance as easy conditions subtracted from hard conditions, univariate ANOVAs were corrected using Hyun-Feldt's correction when reported in the present study. Wilk's Lambda is the chosen statistic used to calculate multivariate significance because it is the exact *F*-statistic and provides better power given that I am looking at effects falling on a single dimension.

4.1 Correlations

I hypothesized that pupil size as calculated using the PDR during the cue or during the end of trial would correlate with measures of cognitive ability. While my cognitive ability measures did correlate with accuracy (see Table 1), none of them correlated with pupil size in either the Garavan or the visual search task. The largest correlation with pupil size in the Garavan task was with the automated operation span task during difficulty cues that directly followed hard trials that were answered correctly, $r(49) = -0.17$. The largest correlation with pupil size in the visual search task

was with Need for Cognition in the linger analysis during difficulty cues that directly followed easy trials that were answered correctly, $r(85) = 0.20$. Due to the lack of relevant correlations, automated operation span, Raven's Progressive Matrices, and Need for Cognition measures were not included as covariates in further analyses.

Table 1 -- Correlations between cognitive ability variables and accuracy.

	1	2	3	4	5	6	7	8	9	10	N	Mean	SD
1. Need for Cognition Scale	-										95	19.25	18.49
2. Automated Operation Span Task	-0.01	-									94	63.15	11.16
3. Raven's Progressive Matrices Odd Proble	0.12	0.26*	-								95	11.90	2.82
4. Composite Z-score for Cognitive Ability	0.07	0.79**	0.80**	-							94	0.10	1.61
5. Garavan Accuracy for Easy Trials	0.04	0.28*	0.34*	0.39**	-						56	0.65	0.19
6. Garavan Accuracy for Medium Trials	0.09	0.40**	0.24	0.41**	0.70**	-					56	0.37	0.19
7. Garavan Accuracy for Hard Trials	0.07	0.21	0.28*	0.32**	0.49**	0.65**	-				56	0.17	0.14
8. Visual Search Accuracy for Easy Trials	-0.23	0.00	-0.02	-0.01	0.25*	0.13	0.11	-			88	0.83	0.06
9. Visual Search Accuracy for Medium Trial	-0.08	0.12	-0.02	0.06	0.23*	0.26*	0.24*	0.32**	-		88	0.34	0.11
10. Visual Search Accuracy for Hard Trials	0.00	0.05	0.01	0.04	0.19	0.25*	0.26*	0.21*	0.73**	-	88	0.27	0.11

Notes. * = significant at $p < 0.05$ level. ** = significant at $p < 0.01$ level.

4.2 Garavan Full Analyses

Before running analyses on the Garavan task, I filtered out participants who had not scored at least one point for each level of difficulty in the task. I was left with a total of 56 participants. I then ran a 3 (measure: accuracy, pupil size during difficulty cue, pupil size at end of trial fixation) x 3 (difficulty: easy, medium, hard) totally within-subjects doubly multivariate MANOVA. Accuracy in the present context is percentage correct.

The multivariate effect of difficulty was significant; one of my measures varied depending on the different levels of difficulty, $F(6, 216) = 47.99$, $p < 0.01$. I ran univariate analyses as well as paired *t*-tests to determine where effects were occurring.

Paired *t*-tests were corrected using Bonferroni's technique to account for increases in family wise error. All tests are significant at $p < 0.05$.

The univariate effect of accuracy was significant meaning that accuracy varied depending on level of difficulty, $F(2, 109.95) = 221.65$, $p < 0.01$. I ran three paired *t*-tests to understand this effect. Accuracy for easy trials ($M = 0.65$, $SD = 0.19$) was significantly higher than accuracy for medium trials ($M = 0.37$, $SD = 0.19$; $t(55) = 12.26$, $p < 0.05$) and hard trials ($M = 0.17$, $SD = 0.14$; $t(55) = 19.60$, $p < 0.05$). Accuracy for medium trials was significantly higher than accuracy for hard trials, $t(55) = 9.29$, $p < 0.05$.

Pupil size during the difficulty cue was significant at the univariate level meaning that pupil size during the difficulty cue depended on the level of difficulty, $F(1.93, 106.38) = 4.17$, $p < 0.05$. I ran three paired *t*-tests to understand this effect. Pupil size during easy cues ($M = 1.08$, $SD = 0.22$) was not significantly different from pupil size during medium cues ($M = 1.08$, $SD = 0.22$; $t(55) = 1.61$, $p > 0.05$) or hard cues ($M = 1.09$, $SD = 0.23$; $t(55)=1.18$, $p > 0.05$). Pupil size during medium cues was significantly smaller than pupil size during hard difficulty cues, $t(55) = 3.21$, $p< 0.05$.

Pupil size during the end of trial fixation was significant at the univariate level meaning that pupil size at the end of trial was dependent on level of difficulty, $F(1.91, 104.90) = 7.99$, $p < 0.01$. I ran three paired *t*-tests to understand this effect. Pupil size at the end of hard trials ($M = 1.01$, $SD = 0.23$) was significantly larger than pupil size at the end of easy ($M = 0.99$, $SD = 0.22$; $t(55) = 3.55$, $p < 0.05$) or medium trials ($M = 1.00$, $SD = 0.23$; $t(55) = 3.16$, $p < 0.05$). Pupil size at the end of easy trials was not significantly different from pupil size at the end of medium trials, $t(55) = 1.18$, $p >0.05$.

4.3 Garavan Accuracy Analyses

To find out if pupil size was dependent on performance, I compared pupil size for correct trials to pupil size for incorrect trials. I ran a new 3(difficulty: easy, medium, hard) x 2(Accuracy: correct trials, incorrect trials) x 2(measure: pupil size during difficulty cue, pupil size at end of trial fixation) totally within-subjects doubly multivariate MANOVA. If the effect of accuracy or the interaction of accuracy and difficulty were significant, then performance had an impact on pupil size. I only included participants who had at least one right answer and at least one wrong answer for each level of difficulty to ensure that I was comparing the same people across difficulty levels – that is, to ensure that people who could answer even the most difficult trials correctly were not being compared with people who could not answer any of the difficult trials correctly. I was left with 55 participants.

Neither the multivariate effect of accuracy ($F(2, 53) = 0.25$; $p > 0.05$) nor the interaction of accuracy and difficulty ($F(4, 216) = 1.78$; $p > 0.05$) were significant. The multivariate effect of difficulty was significant – an expected finding based on the previous analysis, $F(4, 214) = 2.64$, $p < 0.05$. Unlike the previous analysis, at the univariate level, only pupil size at the end of trial was different depending on difficulty, $F(1.90, 102.37) = 3.75$; $p < 0.05$. Because marginal means for the main effect of difficulty in this condition are, by necessity, identical to marginal means for the main effect of difficulty in the previous analysis, paired-sample *t*-tests were not run. Means are reported in Table 2.

Table 2 -- Means for the Garavan accuracy condition.

Measure	Accuracy	Difficulty	Mean	SD
PDR during cue presentation	Incorrect	Easy	1.1	0.28
		Medium	1.07	0.23
		Hard	1.09	0.24
	Correct	Easy	1.09	0.22
		Medium	1.09	0.22
		Hard	1.1	0.23
PDR at end of trial fixation	Incorrect	Easy	1	0.23
		Medium	0.99	0.24
		Hard	1.01	0.24
	Correct	Easy	0.98	0.22
		Medium	1.01	0.21
		Hard	1.02	0.24

Notes. N = 55.

4.4. Garavan Linger Analyses

To see if previous trial performance or difficulty could change pupil size during the difficulty cue for the present trial, I ran a 3(difficulty for the previous trial: easy, medium, hard) x 2(Accuracy for the previous trial: correct, incorrect) x 1(pupil size during the difficulty cue for the present trial) GLM repeated measures ANOVA.

Neither of the main effects nor the interaction were significant (see Table 3). This suggests that pupil size during the difficulty cue did not change depending on previous trial performance or difficulty of the previous trial.

Table 3 -- Univariate results for the Garavan linger analyses.

Effect	Hypothesis df	Error df	*F* -value	Significance
Difficulty	1.64	82.02	2.51	$p > 0.05$
Accuracy	1.00	50.00	0.07	$p > 0.05$
Interaction	1.91	95.67	2.16	$p > 0.05$

4.5 Visual Search Full Analyses

Before running analyses on the visual search task, I filtered out people who did not have at least one accurate response and one inaccurate response for each level of difficulty. I was left with 88 people. I then ran a 3 (measure: accuracy, pupil size during difficulty cue, pupil size at end of trial fixation) x 3 (difficulty: easy, medium, hard) totally within subjects doubly multivariate MANOVA. Accuracy in the present context is percentage correct.

The multivariate effect of difficulty was significant; one of my measures varied depending on the different levels of difficulty, $F(6, 344) = 226.44$, $p < 0.01$. I ran univariate analyses as well as paired *t*-tests to determine where effects were occurring. Paired *t*-tests were corrected using Bonferroni's technique to account for increases in family wise error. All tests are significant at $p < 0.05$.

The univariate effect of accuracy was significant, meaning that accuracy varied depending on level of difficulty, $F(1.78, 154.80) = 1565.11$, $p < 0.01$. I ran three paired *t*-tests to understand this effect. Accuracy for easy trials ($M = 0.83$, $SD = 0.06$) was significantly higher than accuracy for medium trials ($M = 0.34$, $SD = 0.11$; $t(87) = 41.97$, $p < 0.05$) or hard trials ($M = 0.27$, $SD = 0.11$; $t(87) = 46.30$, $p < 0.05$). Accuracy

for medium trials was significantly higher than accuracy for hard trials, $t(87) = 7.82$, $p < 0.05$.

Pupil size during the difficulty cue was significant at the univariate level, meaning that pupil size during the difficulty cue depended on the level of difficulty, $F(1.44, 124.83) = 6.46$, $p < 0.05$. I ran three paired *t*-tests to understand this effect. Pupil size during easy cues ($M = 0.97$, $SD = 0.18$) was not significantly different from pupil size during medium cues, $M = 0.97$, $SD = 0.16$; $t(87) = 1.23$; $p > 0.05$. Pupil size during hard cues ($M = 0.98$, $SD = 0.18$) was significantly larger than pupil size during easy cues ($t(87)=3.53$, $p < 0.05$) and medium cues, $t(87) = 2.87$; $p < 0.05$.

Pupil size during the end of trial fixation was significant at the univariate level meaning that pupil size at the end of trial was dependent on level of difficulty, $F(1.72, 149.50) = 122.67$, $p < 0.01$. I ran three paired *t*-tests to understand this effect. Pupil size at the end of easy trials ($M = 0.92$, $SD = 0.17$) was significantly larger than pupil size at the end of medium trials ($M = 0.87$, $SD = 0.15$; $t(87) = 9.41$; $p < 0.05$) and hard trials, $M = 0.84$, $SD = 0.16$; $t(87) = 13.70$; $p < 0.05$. Pupil size at the end of medium trials was significantly larger than pupil size at the end of hard trials, $t(87) = 6.97$; $p < 0.05$.

4.6 Visual Search Complete Accuracy Analyses

People were instructed to search for the letter "P" hidden in an array of distracter letters. Instructions informed participants that there was a letter "P" in every trial and that they had ten seconds to respond before the trial would end. Because of the nature of my instructions, the visual search task inadvertently provided people with performance feedback; if people did not make a response before time ran out (miss trials) or clicked on anything other than the letter "P" (false-alarm trials), then the trial

would be scored as incorrect. This setup allows me to investigate the relationship between trial difficulty, performance feedback, emotional arousal, and attention. To test these relationships, I ran a 3 (difficulty: easy, medium, hard) x 3 (accuracy: hit, miss – or time out trials, false-alarm – or a trial where a person clicked on something other than the "P") x 2 (recording point: pupil size during difficulty cue, pupil size at the end of trial fixation) totally within subjects doubly multivariate GLM MANOVA. Similar to previous analyses, pupil size during the difficulty cue is thought to represent preparatory attention processes and/or emotional arousal evoked by cue content. Unlike previous analyses, pupil size at the end of trial fixation is thought to represent either attention processes that may still be online from the previous trial (especially likely for miss trials where the trial ends before a person is done searching) or emotional arousal evoked through performance feedback, or both. Before beginning analyses, I filtered out participants who did not have at least one hit, false-alarm, and miss for each level of difficulty. I was left with 67 participants.

The main effect of difficulty was significant at the multivariate level, $F(4, 262) = 22.41$; $p < 0.05$. The main effect of accuracy was significant at the multivariate level, $F(4, 262) = 7.60$; $p < 0.05$. Because the interaction between difficulty and accuracy was significant at the multivariate level, the analyses for this condition will focus on the interaction effect, $F(8, 526) = 5.04$; $p < 0.05$.

Pupil size during the difficulty cue did not differ significantly depending on the interaction of accuracy and difficulty, $F(1.16, 76.47) = 0.03$; $p > 0.05$. Pupil size at the end of trial fixation significantly differed depending on the interaction of accuracy and difficulty, $F(1.93, 127.07) = 9.49$; $p < 0.05$.

I ran nine paired-sample *t*-tests examining accuracy conditioned on difficulty. For easy trials, there was no significant difference in end-of-trial pupil size for false-alarm trials ($M = 0.94$, $SD = 0.22$) and hit trials ($M = 0.92$, $SD = 0.17$; $t(66) = 1.36$; $p > 0.05$); or between false-alarm and miss trials ($M = 0.91$, $SD = 0.21$; $t(66) = 1.25$; $p > 0.05$); or hit trials and miss trials, ($t(66) = 0.53$; $p > 0.05$). For medium trials, there was a significant difference in end-of-trial pupil size for false-alarm trials ($M = 0.85$, $SD = 0.17$) and hit trials ($M = 0.91$, $SD = 0.16$; $t(66) = 4.90$; $p < 0.05$); and between hit and miss trials ($M = 0.83$, $SD = 0.16$; $t(66) = 6.97$; $p < 0.05$). There was not a significant difference in end-of-trial pupil size between false-alarm trials and miss trials for medium difficulty conditions, ($t(66) = 1.78$; $p > 0.05$). For hard trials, there was a significant difference in end-of-trial pupil size for false-alarm trials ($M = 0.82$, $SD = 0.16$) and hit trials ($M = 0.90$, $SD = 0.17$; $t(66) = 6.56$; $p < 0.05$); and between hit and miss trials ($M = 0.81$, $SD = 0.18$; $t(66) = 5.29$; $p < 0.05$). There was not a significant difference in end-of-trial pupil size between false-alarm trials and miss trials for medium difficulty conditions, ($t(66) = 0.83$; $p > 0.05$).

I ran nine paired-sample *t*-tests examining difficulty conditioned on accuracy. For false-alarm trials, end-of-trial pupil size was significantly larger for easy trials than for medium trials ($t(66) = 8.11$; $p < 0.05$) or hard trials ($t(66) = 8.06$; $p < 0.05$); and end-of-trial pupil size was significantly larger for medium trials than for hard trials, $t(66) = 4.03$; $p < 0.05$). For hit trials, easy trials were not significantly different from medium trials ($t(66) = 1.29$; $p > 0.05$) or hard trials ($t(66) = 1.54$; $p > 0.05$); nor were medium trials significantly different from hard trials, $t(66) = 0.86$; $p > 0.05$. For miss trials, end-of-trial pupil size was significantly larger for easy trials than for medium

trials ($t(66) = 3.51$; $p < 0.05$) and hard trials ($t(66) = 4.46$; $p < 0.05$), however, end-of-trial pupil size was not significantly different between hard trials and medium miss trials, $t(66) = 2.69$; $p > 0.05$.

4.7 Visual Search Linger Analyses

To see if previous trial performance or difficulty could change pupil size during the difficulty cue for the present trial, I ran a 3(difficulty for the previous trial: easy, medium, hard) x 2(Accuracy for the previous trial: correct, incorrect) x 1(pupil size during the difficulty cue for the present trial) GLM repeated measures ANOVA.

The main effects of accuracy ($F(1, 87) = 2.23$; $p > 0.05$) and difficulty ($F(1.24, 107.80) = 0.93$; $p > 0.05$) were not significant. The interaction of accuracy and difficulty were not significant, $F(1.72, 149.60) = 1.98$, $p > 0.05$. This suggests that neither performance nor difficulty for the previous trial significantly impacted pupil size during the difficulty cue.

CHAPTER 5. DISCUSSION

There is considerable evidence suggesting a link between perceived task difficulty and pupil size with larger pupil sizes associated with more difficult tasks (Jiang, Zheng, Bednarik, Atkins, 2015; Nuthmann & van der Meer, 2005; Raisig, Welke, Hagendorf, & van der Meer, 2007; Verney, Granholm, & Marshall, 2004). Additionally, there is a relationship between pupil size and emotion with pupil size being larger for more emotional stimuli regardless of valence (Bradley, Miccoli, Escrig, & Lang, 2008; Partala & Surakka, 2003). In the present study, I tested the extent to which preparatory attention processes were engaged by a basic difficulty cue (i.e., easy, medium, hard) using two conceptually different tasks: a sound-based Garavan task and a visual search task. The Garavan task measures memory updating (Garavan, 1998) and correlates with cognitive ability in the present study. The visual search task still requires attention control to maintain task instructions and to perform efficient searches; however, attention control strategies are more implicit within the visual search task than the Garavan task. Unlike the Garavan task, performance on the visual search task could be inferred because people (1) were informed that there was a target in each trial and (2) had to click on the target before the trial timed out in order for the trial to be scored as correct.

I presented people with valid difficulty cues before each trial for both tasks and recorded pupil size during the viewing of the cue as well as at an end of trial fixation. To tease apart changes in pupil size induced by emotional arousal (e.g., anxiety during the viewing of the cue for more difficult trials that may not be present in easier trials)

from what could be preparatory attention control, I included accuracy as an independent variable and compared pupil size across difficulty levels and between correct and incorrect trials. While the interaction between cue and subsequent accuracy should reveal the extent to which preparatory attention processes were sufficient to improve performance, the main effect of difficulty should primarily reveal information about emotional processing. Because the eye-tracker reported pupil size in arbitrary units, I inferred pupil size using a pupillary dilation ratio (see: Pupillary Data Filtering).

I had originally hypothesized that measures of cognitive ability – that is, Raven's Progressive Matrices and the Automated Operation Span Task – as well as measures of mental engagement – that is, the Need For Cognition task – would correlate with pupil size as well as task performance. Performance on the sound-based Garavan task correlated with both measures of cognitive ability but failed to correlate with cognitive engagement. Visual search task performance related to neither cognitive ability nor cognitive engagement in the present study. There was no relationship between pupil size during the difficulty cue and during the end of trial fixation for either cognitive ability or cognitive engagement in the present study.

5.1 Findings for the Sound-Based Garavan

Accuracy decreased as difficulty increased suggesting (1) easy trials were easier than both medium and hard trials and (2) medium trials were easier than hard trials. Yet, there is reason to believe that my labels may not have been accurate. Trials labeled "easy" were answered correctly 65% of the time – performance conceptually more similar to a medium difficulty level. Additionally, trials labeled "medium" were answered correctly 37% of the time – performance conceptually equivalent to hard

difficulty levels. Trials labeled as “hard” were answered correctly 17% of the time suggesting that these trials were too hard.

When accuracy was included as an independent variable, neither accuracy nor the interaction between cue/difficulty and accuracy were significant. The main effect of difficulty was only significant for pupil size at the end-of-trial fixation when emotional arousal is likely to have greater influence on pupil size. Pupil size at the end of hard trials was significantly larger than pupil size at the end of easy and medium trials. There are two non-mutually exclusive explanations. First, it is possible that hard trials required greater attention control causing an increase in pupil size during the trial. At the end of the trial, pupil size would be larger for hard trials than for other difficulties and may take longer to return to baseline. Second, it is possible that there was more emotional arousal at the end of hard trials than at the end of easy or medium trials. Because end-of-trial pupil size was not significantly different between easy and medium trials, it is difficult to distinguish between these two possibilities.

When accuracy was included as a dependent variable, pupil size during the cue was significantly different across difficulty levels. Pupil size was significantly smaller for medium cues than hard cues; however, pupil size during easy cues was not significantly different than medium or hard trials. Because pupil size during the cue did not vary when including accuracy as an independent variable, I believe this finding may not be reliable. Regardless, pupil size during the cue was not influenced by previous trial accuracy or difficulty suggesting that end of trial pupil size had returned to a sufficient baseline at the start of the next trial.

Despite very low accuracy in hard conditions, it does not appear that people were giving up. If people were giving up on the trials, I would not expect end-of-trial pupil size for the hardest trials to be significantly larger than easy or medium trials.

5.2 Findings for the Visual Search Task

Like the Garavan task, accuracy decreased as difficulty increased suggesting (1) easy trials were easier than both medium and hard trials and (2) medium trials were easier than hard trials. Yet, there is reason to believe that my labels may not have been accurate. Trials labeled "easy" were answered correctly 83% of the time which is performance conceptually expected for easy conditions. Based on performance, trials labeled "medium" (34% accuracy) and trials labeled "hard" (27% accuracy) were not that conceptually different from hard trials.

Additionally, pupil size was consistently lower than baseline in this task. The Garavan task lasted at least 45 minutes and the visual search task was always presented after. There is a continuous decrease in pupil size until the end of the experiment (Kahneman & Peavler, 1969). Additionally, fatigue causes a decrease in pupil size (Lowenstein & Loewenfeld, 1964). For these reasons, I do not find the fact that pupil size was below baseline to be either novel or task related.

Pupil size during the difficulty cue depended on cue content when subsequent accuracy was included as a dependent variable. Pupil size was not different between easy and medium cues. Pupil size during easy and medium cues was significantly smaller than pupil size during hard cues. Similar to the Garavan task, this finding does not appear to be robust since no effects of pupil size during the cue were significant when subsequent accuracy was included as an independent variable. For this reason, it

seems prudent not to interpret these findings. Regardless, pupil size during the difficulty cue was not influenced by previous trial accuracy or difficulty.

For end-of-trial pupil size, the interaction between difficulty and accuracy was significant. Thus, pupil size may reflect both emotional arousal from inferring task performance and attention control processes that may be briefly active at the end of the trial. Interpretations depend on whether or not people were able to infer task performance. If people were able to infer task performance, emotional arousal likely had a greater influence on end-of-trial pupil size.

5.3 Could People Infer Task Performance In The Visual Search Task?

I believe that people were able infer task performance accurately in the visual search task for several reasons. First, pupil size was the same for all accuracy types at the end of easy trials. It is unlikely that people cared less about answering correctly or incorrectly in easy trials considering that (1) pupil size for hit trials was not significantly different at the end of easy, medium, and hard trials (2) pupil size for hit trials was significantly larger than false alarm or miss trials in medium and hard conditions, and (3) false alarms and misses were significantly larger in easy conditions than in medium or hard conditions.

Second, pupil size for misses and false alarms were not significantly different at the end of easy, medium, and hard trials. Thus, people were experiencing similar emotional arousal and/or engaged similar amounts of attention control evoking statistically identical pupil sizes for both false alarms and misses. I believe the only way this pattern could occur is if people were aware of their trial performance.

Finally, previous findings show false alarms evoke pupil sizes at least as large as correct trials when people were not aware of having made a false alarm (Montefinese, Ambrosini, Fairfield, Mammarella, 2013; Otero, Weekes, Hutton, 2011). My findings show false alarms evoking similar pupil size to miss trials in medium and hard conditions. Additionally, false alarms were smaller than hit trials in medium and hard conditions. Thus, I am confident that people were able to infer task performance in my visual search task.

5.4 Is There More Evidence For Emotional Or Attentional Explanations For Differences In Pupil Size?

I believe the differences in end of trial pupil size are caused by emotional arousal rather than by attention control for several reasons. First, despite increases in difficulty, there were no significant differences in end-of-trial pupil size for hit trials. It is possible that visual search tasks require a threshold level of attention control that, once met, trials will be answered correctly regardless of difficulty – after all, hit trials were significantly larger than false alarms and misses in medium and difficult conditions. When considering that answering trials correctly was a relatively infrequent event, it is also plausible that people experienced more emotional arousal (e.g., excitement) for answering correctly and this emotional arousal was equivalent across difficulty levels.

Second, there was no difference in end-of-trial pupil size between false alarm, hit, and miss trials in easy conditions *and* end-of-trial pupil size was larger for easy trials than medium or hard trials for false alarms and misses. Additionally, false alarms evoke larger end-of-trial pupil size in medium than hard conditions. This fits what

would be expected with emotional arousal – in easy conditions, there is more emotional arousal (e.g., frustration or surprise) for answering incorrectly because there is an expectation that easy trials will be answered correctly; the target stands out and is usually locatable. For medium and hard conditions, misses may be more commonplace evoking less emotional arousal. It is also possible that people were more likely to give up in medium and hard conditions – either clicking a non-target to advance the trial or waiting passively until the trial timed out.

5.5 Limitations

This study was not without limitations. First, if the interaction between accuracy and difficulty are not significant, it is next to impossible to test if changes in pupil size are due to differences in attention control or emotional arousal using only pupillometry. Furthermore, pupillometry may not be the best method to see differences in preparatory attention control. Future studies would benefit from pairing pupillometry with neuroimaging techniques like electroencephalography and event-related potential and functional magnetic resonance imaging to couple changes in pupil size with corresponding changes in neural activity in related brain regions.

Second, the nature of my sample may explain why pupil size did not correlate with measures of cognitive ability. Participants were limited to students at a prestigious university and are unlikely to be representative of the general population. Future studies would benefit from a more diverse population.

Third, I did not ask people what their motivation level was while they were completing the task. Future studies would benefit from collecting trial relevant

motivational information to better tease apart attention control and emotional arousal from the pupillary response.

Fourth, it is possible that I did not find differences in pupil size during the difficulty cues because my cues were too rudimentary. Using cues that are more informative may help people better prepare to meet task demands and may raise preparatory attention, if it is occurring, to a sufficient level to impact performance. Additionally, it may be useful to consider using a task where a person may be able explicitly prepare different strategies such as a stroop or visual search task rather than a Garavan task where preparation for an upcoming trial is implicit.

Finally, while accuracy decreased significantly as difficulty increased, my trials do not appear to be appropriately labeled for the conceptual difficulty level of the trials. Future studies would benefit from better norming so that easy trials had 85% accuracy, medium trials had 65% accuracy, and hard trials had accuracy below 50%. People may have relied less upon the difficulty cue because labeled accuracy was not conceptually matched with difficulty.

REFERENCES

Aboyoun, D. C., & Dabbs, J. M. (1998). The Hess pupil dilation findings: Sex or novelty?. *Social Behavior and Personality: an international journal*, *26*(4), 415-419.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci., 28*, 403-450.

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *The Journal of Neuroscience, 14*(7), 4467-4480.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology*, *2*, 142-162.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602-607.

Bradshaw, J. L. (1968). Pupil size and problem solving. *The Quarterly journal of experimental psychology*, *20*(2), 116-122.

Bumke, O. (1896). translated by Eckhard Hess in The Tell Tale Eye (1975), pp. 23–24. New York: Van Nostrand. 1975, 1911.

Cacioppo, J. T., Petty, R. E., Feinstein, J, A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197–253.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.

Ester, E. F., Vogel, E. K., & Awh, E. (2012). Discrete resource limits in attention and working memory. In M. I. Posner (Ed.) , *Cognitive neuroscience of attention (2nd ed.)* (pp. 99-110). New York, NY, US: Guilford Press.

Franco, A. (2010). Resource allocation of the human brain: A competitive equilibrium approach. *Dissertation Abstracts International*, *71*, 1144.

Garavan, H. (1998). Serial attention within working memory. *Memory & cognition, 26*(2), 263-276.

Gilzenrat, M. S., Cohen, J. D., Rajkowski, J., & Aston-Jones, G. (2003). Pupil dynamics predict changes in task engagement mediated by locus coeruleus. In *Society for Neuroscience Abstracts* (Vol. 515, p. 19)

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

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457-461.

Heaver, B. (2012). *Psychophysiological indices of recognition memory* (Doctoral dissertation, University of Sussex).

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

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190-1192.

Heinrich, W. (1896) Die Aufmerksamkeit und die Funktion der Sinnesorgane. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 9*, 342-388.

Janisse, M. P. (1974). Pupil size, affect and exposure frequency. *Social Behavior and personality*, *2*(2), 125-146.

Jiang, X., Zheng, B., Bednarik, R., & Atkins, M. S. (2015). Pupil responses to continuous aiming movements. *International Journal of Human-Computer Studies*, *83*, 1-11.

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. *Journal of cognitive neuroscience*, *23*(7), 1587-1596.

Jiang, X., Zheng, B., Bednarik, R., & Atkins, M. S. (2015). Pupil responses to continuous aiming movements. *International Journal of Human-Computer Studies*, *83*, 1-11.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 310.

Beatty, J., & Kahneman, D. (1966a). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371-372.

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

Kahneman, D., & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, *2*(3), 101-105.

Kahneman, D., & Peavler, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology*, *79*(2p1), 312.

Kuchinke, L., Võ, M. L. H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132-140.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23-34.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry a window to the preconscious?. *Perspectives on psychological science*, *7*(1), 18-27.

Lange, K., & Schnuerch, R. (2014). Challenging perceptual tasks require more attention: The influence of task difficulty on the N1 effect of temporal orienting. *Brain and cognition*, *84*(1), 153-163.

Lewis, D. I., & Coote, J. H. (1990). Excitation and inhibition of rat sympathetic preganglionic neurones by catecholamines. Brain research, 530(2), 229-234.

Libby, W. L., Lacey, B. C., & Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology*, *10*(3), 270-294.

Loewenfeld, I. E. (1993). *The pupil: Anatomy, physiology, and clinical applications*. Ames: Iowa State University Press.

Lowenstein O., Loewenfeld I.E. (1962). *The pupil*. In: H Davidson, editor. The eye. New York: Academic Press. pp. 255–340.

Lowenstein, O., & Loewenfeld, I. E. (1964). The sleep-waking cycle and pupillary activity. *Annals of the New York Academy of Sciences*, *117*(1), 142-156.

Maw, N. N., & Pomplun, M. (2004). Studying human face recognition with the gaze-contingent window technique. In *Proceedings of the Twenty-Sixth Annual*

Meeting of the Cognitive Science Society, 2004, Chicago, IL (pp. 927-932). Mahwah, NJ: Erlbaum.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The "subjective" pupil old/new effect: is the truth plain to see?. *International Journal of Psychophysiology*, *89*(1), 48-56.

Moresi, S., Adam, J. J., Rijcken, J., Van Gerven, P. W., Kuipers, H., & Jolles, J. (2008a). Pupil dilation in response preparation. *International Journal of Psychophysiology*, *67*(2), 124-130.

Moresi, S., Adam, J. J., Rijcken, J., & Van Gerven, P. W. (2008b). Cue validity effects in response preparation: a pupillometric study. *Brain research*, *1196*, 94-102.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532-1543.

Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, *42*(3), 306-317.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 1346-1353.

Partala, T., Jokiniemi, M., & Surakka, V. (2000, November). Pupillary responses to emotionally provocative stimuli. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 123-129).

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, *59*(1), 185-198.

Payne, D. T., Parry, M. E., & Harasymiw, S. J. (1968). Percentage of pupillary dilation as a measure of item difficulty. *Perception & Psychophysics*, *4*(3), 139-143.

Peavler, W. S. (1974). Pupil size information overload, and performance differences. *Psychophysiology*, *11*(5), 559-566.

Phillips, M. A., Szabadi, E., & Bradshaw, C. M. (2000). Comparison of the effects of clonidine and yohimbine on spontaneous pupillary fluctuations in healthy human volunteers. *Psychopharmacology, 150*(1), 85-89.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology*, *60*(2), 211-229.

Raisig, S., Welke, T., Hagendorf, H., & Meer, E. V. D. (2007). Investigating dimensional organization in scripts using the pupillary response. *Psychophysiology, 44*(6), 864-873.

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. In *Society for Neuroscience Abstracts* (Vol. 19, p. 974).

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current neuropharmacology, 6*(3), 254.

Schiff, J.M. (1875). *La pupille consideré comme esthésiom ètre* (translated by R.G. de Choisity). Paris: J.B. Balliére.

Slagter, H. A., Prinssen, S., Reteig, L. C., & Mazaheri, A. (2016). Facilitation and inhibition in attention: Functional dissociation of pre-stimulus alpha activity, P1, and N1 components. *NeuroImage*, *125*, 25-35.

Smith, M. S., Schambra, U. B., Page, S. O., Hulette, C., & Schwinn, D. A. (1994). [alpha] 2-Adrenergic Receptors in Human Spinal Cord: Specific Localized Expression of mRNA Encoding [alpha] 2-Adrenergic Receptor Subtypes at Four Distinct Levels. *Anesthesiology*, 81(3A), A1492.

Stanners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion, 3*(4), 319-340.

Steinhauer, S. R., Boller, F., Zubin, J., & Pearlman, S. (1983). Pupillary dilation to emotional visual-stimuli revisited. *Psychophysiology, 20*(4), 472-472.

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International journal of psychophysiology, 52*(1), 77-86.

Tucker, A., & Johns, M. W. (2005, June). The duration of eyelid movements during blinks: changes with drowsiness. Poster presented at the 19th annual Associated Professional Sleep Societies, LLC "Sleep" conference, Denver, CO.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior research methods*,*37*(3), 498-505.

van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PLoS One*, *7*(12), e51134.

Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, *52*(1), 23-36.

Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, *45*(1), 130-140.

Woodmansee, J. J. (1966). Methodological Problems in Pupillographic Experiments. In *Proceedings of the Annual Convention of the American Psychological Association*. American Psychological Association.