Old Dominion University
ODU Digital Commons

Psychology Theses & Dissertations | Psychology

Summer 2021

Predicting Inattentional Blindness with Pupillary Response in a Simulated Flight Task

Kellie D. Kennedy
Old Dominion University, kellie.d.kennedy@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_etds

Part of the Biological Psychology Commons, Cognitive Psychology Commons, Health Psychology Commons, and the Human Factors Psychology Commons

Recommended Citation
Kennedy, Kellie D.. "Predicting Inattentional Blindness with Pupillary Response in a Simulated Flight Task" (2021). Doctor of Philosophy (PhD), Dissertation, Psychology, Old Dominion University, DOI: 10.25777/y9qm-8551
https://digitalcommons.odu.edu/psychology_etds/369

This Dissertation is brought to you for free and open access by the Psychology at ODU Digital Commons. It has been accepted for inclusion in Psychology Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

PREDICTING INATTENTIONAL BLINDNESS WITH PUPILLARY RESPONSE IN A SIMULATED FLIGHT TASK

by

Kellie D. Kennedy
B.S. May 2009, Wright State University
M.S. May 2012, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

PSYCHOLOGY

OLD DOMINION UNIVERSITY
August 2021

Approved by:

Michelle L. Kelley (Director)

James R. Comstock (Member)

Alan T. Pope (Member)

Abby L. Braitman (Member)

ABSTRACT

PREDICTING INATTENTIONAL BLINDNESS WITH PUPILLARY RESPONSE IN A SIMULATED FLIGHT TASK

Kellie D. Kennedy
Old Dominion University, 2021
Director: Dr. Michelle Kelley

Inattentional blindness (IB) is the failure of observers to notice the presence of a clearly viewable but unexpected visual event when attentional resources are diverted elsewhere. Knowing when an operator is unable to respond or detect an unexpected event may help improve safety during task performance. Unfortunately, it is difficult to predict when such failures might occur. The current study was a secondary data analysis of data collected in the Human and Autonomous Vehicle Systems Laboratory at NASA Langley Research Center. Specifically, 60 subjects (29 male, with normal or corrected-to-normal vision, mean age of 34.5 years (*SD* = 13.3) were randomly assigned to one of three automation conditions (full automation, partial automation, and full manual) and took part in a simulated flight landing task. The dependent variable was the detection/non-detection of an IB occurrence (a truck on the landing runway).

Scores on the NASA-TLX workload rating scale varied significantly by automation condition. The full automation condition reported the lowest subjective task load followed by partial automation and then manual condition. IB detection varied significantly across automation condition. The moderate workload condition of partial automation exhibited the lowest likelihood of IB occurrence. The low workload full automation condition did not differ significantly from the manual condition. Subjects who reported higher task demand had increased pupil dilation and subjects with larger pupil dilation were more likely to detect the

runway incursion. These results show eye tracking may be used to identify periods of reduced unexpected visual stimulus detection for possible real-time IB mitigation.

Copyright, 2021, by Kellie D. Kennedy, All Rights Reserved.

This dissertation is dedicated to those who believed in me and offered their strength when mine faltered.

ACKNOWLEDGMENTS

I would like to acknowledge several people who contributed to the successful completion of my dissertation. First, I wish to acknowledge Dr. Gale Dodge and Dr. Debbie Major. I present this achievement as the deepest form of thanks I can offer for defending my opportunity to finish my education. Next, I would like to acknowledge my advisor, Dr. Kelley, who provided true mentorship with her guidance and encouragement. She was free with her time, helpful with her advice, and kept me always moving forward. I want to acknowledge my committee members, Drs. Abby Braitman, Ray Comstock, and Alan Pope. Their thorough and constructive feedback greatly improved the quality of my research and expanded my thinking in many ways. Special thanks to Paul Schutte for providing me a protected space to work and to dream and without whom these data would not have been possible. I am grateful for the mentorship and friendship I have shared over the last several years with Alan, Ray, and Paul. I am better scientist and person for knowing them. Thank you to my numerous ODU and NASA colleagues, and dear ones, past and present who leaned on me as I leaned on them. Thanks to Matt for putting such joy in the journey, semester after semester, and cheering me on through more than a few setbacks.

I would like to thank my family for their constant love and encouragement. Thanks to my parents who taught me the importance of making myself proud and to always remember who I am. Thanks to Chad for inspiring me with the most genuine and steadfast example of what a good scientist, person, and partner looks like and how to ‘do better’ with intention and purpose. And, finally, thank you to our son Lucian and his new sibling on the way for giving me the extra motivation to close this chapter and move fully into the new one we are all writing together!

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

"We pay attention to what we are told to attend to, or what we're looking for, or what we already know...what we see is amazingly limited." — *Daniel Simons (Heffernan, 2011)*

Humans have a limited capacity for processing information. Selective attention enables humans to allocate these limited cognitive resources to process relevant stimuli and ignore irrelevant stimuli (Broadbent, 1958; Kahneman, 1973; Treisman, 1964). However, this process is imperfect, and we can fail to detect important and relevant stimuli. Inattentional blindness (IB) is defined as the failure of an observer to detect a clearly viewable but unexpected visual event due to diverted attentional resources (Hutchinson, 2019; Jensen et al., 2011; Mack & Rock, 1998; Simons & Chabris, 1999). The consequences of an IB occurrence can range from benign to disastrous depending on the importance of the undetected information. Consider the scenario of searching the house extensively for an object, such as car keys or a travel mug, only to find said object in an extremely conspicuous location. Expressions such as "right under my nose," "if it were a snake, it would have bitten me," and "look-but-fail-to-see," all describe this kind of an experience. Another utterance, "I never even saw them" may follow a detection failure with more serious consequences such as in the aftermath of a car accident.

Simons (2000) noted that outside of the laboratory setting, observers must be aware of objects in the environment to make volitional changes in behavior. For example, a driver prepares to make a left turn at an intersection, looks both ways, detects no on-coming traffic, and proceeds into the intersection. To the driver, proceeding forward into the intersection was an appropriate decision. Unfortunately, the driver failed to detect an oncoming motorcyclist and caused a collision. The driver did not detect the oncoming traffic and therefore had no reason to deviate from the planned action of turning onto the roadway. Examples like this demonstrate cases when the failure to perceive critical objects can be deadly. The individual experiencing IB has incomplete information and generates an inaccurate representation of the external world, which limits the ability to aptly assess and execute subsequent decisions. This introduction is designed to provide a basic framework of this study. The next several chapters are dedicated to a more in-depth discussion of the relevant literature.

Researchers have attributed errors and accidents to the IB phenomenon across nearly all task environments: medical (Lum et al., 2005), aviation (Fischer et al., 1980), nautical (Fraher, 2010), and vehicular (Strayer & Drews, 2007; Strayer et al., 2003). Observers can fail to detect the appearance of an unexpected object within their field of view regardless of importance or relevancy to the task (Mack & Rock, 1998; Memmert, 2006; Simons & Chabris, 1999). IB can occur when stimuli are dangerous or highly unusual (Fraher, 2010; Hyman et al., 2010; Simons & Chabris, 1999).

The danger of IB has been well-documented both in the laboratory and in real world operations; however, predicting IB prior to occurrence remains elusive. Identifying a relationship between eye tracking and IB might permit the use of eye tracking as a mitigation method or safety feature for operators during task performance in any environment that includes highly visual performance tasks or monitoring, such as aviation. Typically, measuring the eye gaze fixation point is considered an appropriate method to determine the location of visual attention in a visual field. However, in the case of IB, looking does not equate to seeing. IB researchers report eye fixation as insufficient to differentiate between detectors and non-detectors, finding that the amount of time subjects fixate on the critical stimulus is not significantly different between those who detected and those who do not (Koivisto et al., 2004; Kuhn & Findlay, 2010; Memmert, 2006; Oktay & Cangöz, 2018).

The complicated explanation behind why looking at an object does not necessarily equate to seeing an object may be found in the neurological underpinnings of IB (Dehais et al., 2020). When traditional precursors of degraded task performance are investigated as precursors to IB, a myriad of counter-intuitive and conflicting findings arise (Beanland & Chan 2016; Dehais et al., 2020; Mack & Rock, 1998; Memmert, 2006; Wright et al., 2013). As described by Dehais et al. (2020), IB appears to sit at a neurophysiological intersection of cognitive function limitations and the biological limitations of the visual system. Although the outcome event is the same (an object is undetected), the conflicting findings in the IB literature may actually be related to the different underlying causes for IB. For example, a subject who experiences IB due to a visual system overload has little in common with a pilot landing at a quiet airport who fails to detect a maintenance vehicle sitting on the landing runway with flashing beacons (TAIC, 2010; Wright et al., 2013). Both individuals fail to detect visual information that is relevant, detectable, and within the useful field of view but because the underlying causes are different, the metrics required to measure the neurophysiological state may also be different.

Eye tracking technology offers more information than just eye fixation. Eye tracking can be used to assess human cognitive state through measurements such as eye movement patterns, pupil diameter changes, and eyelid closures (Mohan et al., 2019; Peißl et al., 2018). Pupil dilation is an eye-based cognitive state measurement that can be collected passively and in real-time without task interruption. Increased pupil dilation indicates greater workload (Recarte et al., 2008). Wright et al. (2013) assessed IB using pupil dilation and found no significant relationship between detectors and non-detectors in effort or primary task performance. However, Wright et al. focused on IB induced by increasing visual load using a desktop-based stimulus and a cluttered visual field detection task that was boring, repetitive, and did not allow free eye movement. As previously described a task that elicits IB using a densely packed visual field requires little in the way of higher order cognitive processes. Indeed, the generalizability of the evidence and the metric is related to the task chosen to elicit IB.

Winn et al. (2018) published a comprehensive article on the mechanics of pupillometry data collection and identified numerous studies indicating that task-evoked pupil dilation was not a single simplistic response summed up in effort but is instead reflective of the intersection of many internal processes such as attention, engagement, arousal, anxiety, and effort. Following the neurophysiological basis, an experiment that uses pupillary response as a method to examine cognitive state conditions surrounding IB occurrences during task operation should use an experimental task that engages greater use of neurophysiological components likely to affect the pupil response, such as planning, decision making, and task engagement similar to those found in an operational environment. Therefore, the potential for pupil dilation as a predictor of IB may be better assessed using an engaging task with a naturalistic collection method.

The current study examines the cognitive state conditions associated with an IB occurrence for an operationally relevant unexpected object to increase understanding of the predictors related to visual stimulus detection during complex task performance. The data in this study were obtained as part of a larger study examining the use of a specific type of EEG analysis to detect IB. The larger study included three IB flight simulation runs and two photic stimulation sessions featuring strobing lights. Portions of the larger study data were previously analyzed by the researcher and the results presented at two conferences (Kennedy et al., 2014; 2017). The previous studies examined IB occurrences across automation conditions (Kennedy et al., 2014), and the change in likelihood of IB occurrence across automation conditions with repeated induction of IB (Kennedy et al., 2017). The current examination expanded to include data on eye tracking to determine if those with greater pupil dilation during IB induction were more likely to detect the critical stimulus than subjects with smaller pupil dilation. This study also featured revised IB case inclusion based on additional participant self-report questions on the IB questionnaire. Data for the present study were derived from the first IB flight simulation run which occurred prior to any other variable exposure or strobing lights.

In the present study, IB occurrences were assessed for an unexpected runway incursion during a simulated flight landing task across varied task load of three automation conditions (fully automated, partially automated, manual). The impact of automation condition was explored on subjective workload and subjective workload on IB. Pupil dilation was examined to determine whether pupil dilation is a successful discriminator between those who detected and did not detect the IB occurrence. The following chapters examine prior literature relevant to IB and aviation in greater detail.

CHAPTER II

CHARACTERISTICS OF THE AVIATION OPERATIONAL ENVIRONMENT

"Flying is hours and hours of boredom sprinkled with a few seconds of sheer terror."

— Major Gregory "Pappy" Boyington (*Driskell et al., 2013*)

Major Boyington is just one of the many professional aviators to use this sentiment to describe the periods of underload and overload that occur in the flight deck. After nearly a century of aircraft technology improvements, the modern flight deck has become increasingly automated with systems that require varying levels of human interaction. The pilot operates primarily in phases of flight that are difficult or unsafe to automate, meaning the modern pilot frequently serves as either the final line of defense against error or as a monitor of highly reliable systems for an extended period of time with little aircraft engagement in between. Humans perform poorly in such circumstances; similar underload-rich environments are often used to elicit performance decrement during vigilance studies and similar overload situations are comparable to the stress experienced by surgeons (Neigel et al., 2020; Singh, 2009).

These lows and highs are rife with opportunity for attentional errors. The flight deck is a mixture of automation, settings, and sounds designed to provide information. From the first flight lesson, new pilots are instructed that vision is the most important tool and controlling visual attention is of utmost priority. The periods of highest workload for pilots in normal flight operations are take-off and landing (Wilson, 2002). Prior to and during landing the pilot is engaged in the arrival checklist, the landing checklist, maintaining communication with Air Traffic Control, and monitoring the radio frequency for potentially important information such as weather disturbances or an aircraft go-around in the local airspace. Maintaining awareness to the environment both inside and outside the aircraft can induce a state of high workload filled with task-critical information. Any failure to detect critical events can produce serious consequences in this setting.

Aviation and Automation

Several mitigating strategies have been investigated to solve performance failures in the flight deck. There are improvements to the pilot such as training, checklists, and communication techniques. There are improvements to the aircraft and the task such as automation, alarms, or visual interfaces. However, for the first and last several minutes during take-off and landing, the safety of modern flight still ultimately rests with the human pilots. In this period of high risk and high workload, the human information processing system can fail and the failure of an operator to detect an object or event during periods of high workload is intuitive. Less intuitive, however, is the failure of a human to detect an object or event in during a period of low workload.

One popular method to reduce the likelihood for cognitive overload is through the use of automation. The goal is for the automation to relieve some of the task performance burden so the human operator can contribute more cognitive resources towards other elements of the task and to monitor the system. Unfortunately, the appropriate use of automation requires insight into the human state. The misapplication of automation can sometimes induce a worse performance outcome such as is the case in skill loss due to reduced task performance frequency or the failed detection of critical information due to automation complacency (Lee & See, 2004; Parasuraman & Riley, 1997; Scerbo et al., 2001).

One method for detecting cognitive state is eye tracking. Di Nocera et al. (2007) examined eye behaviors in the flight deck and determined that eye fixations correlated with reduced workload of phases of flight such as cruise and suggest that eye tracking may provide valuable insight into operator state. However, eye fixations do not equate to seeing. In a review of the literature on the impacts of automation on human complacency and bias, Parasuraman and Manzey (2010) noted that despite operator eye fixations containing salient and critical information about the automated task, this information may go undetected (IB) if operators do not allocate visual attention to the information. Specifically, they reported that attention is crucial for monitoring automated tasks and found a strong relationship between attention and the automation-induced phenomenon of automation complacency. Automation complacency refers to the degraded ability of a human to detect system malfunctions when a system is controlled by automation as compared to the system under manual control (Parasuraman & Manzey, 2010). This complacency describes the decline in performance that occurs when individuals shift from performing a task themselves to monitoring its automation (Bailey & Scerbo, 2007). Across the documented examples of in-task automation complacency, Parasuraman and Manzey (2010) identified task load, system reliability, and known failure rate were all associated with automation complacency. Bailey and Scerbo reported that highly reliable complex systems with infrequent and unexpected problems elicit reduced operator awareness of system states, particularly while monitoring as opposed to directly engaging with the system. Parasuraman and Manzey (2010) concluded that automation complacency most frequently occurs when an operator has a highly reliable automation system coupled with a high, multiple-task load, which easily describes the modern flight deck in the “hours and hours of boredom” that Major Boyington described. Indeed, Parasuraman and Manzey examined numerous aviation safety reports while defining automation complacency.

Regardless of high or low cognitive workload, the key feature of an IB occurrence is that an object or event is not detected. Although the outcome is the same, the neuropsychological underpinnings (attention limitation or visual processing limitation) related to the detection failure are different and require different mitigation strategies. The current standard of IB assessment is post-experimental retrospective report as documented by Mack and Rock (1998). Unfortunately, this method cannot mitigate or prevent the damaging impact of IB.

Awareness of the human cognitive state in the minutes leading up to landing could provide an invaluable addition to safety if that automation system could predict compromised visual detection ability and engage a mitigation strategy (Fairclough et al., 2013; Pope et al., 2014; Scerbo et al., 2003; Stephens et al., 2018). Greater awareness via passive observation systems can provide information regarding the human state of awareness that might hope to prevent an accident, an unnecessary alarm, or deploy an emergency procedure. In-task identification of sub-optimal human state would provide valuable insight and enable the deployment of potential mitigation strategies. For example, when flying, the co-pilot can easily see the nonverbal signs of a pilot that is at risk of falling asleep. The co-pilot can use this information to re-engage the pilot or even suggest taking over control of the aircraft. Automation, however, is currently blind to such information. The only awareness of the pilot available to the automation is when the pilot interacts with the system. In the event of an inattentive pilot, that interaction may be inaccurate, occur too late, or not happen at all. Digitizing signals associated with inattention could provide increased safety assurance through the engagement of advanced mitigations or interventions prior to an unwanted outcome.

Runway Incursions

The IB task chosen for the current study was a simulated runway incursion task. This type of task was chosen because many elements of modern flight have been automated; however, landing remains in the hands of the pilot. During landing, the pilots make the final determination that the aircraft will continue to touchdown. This decision requires both the aircraft to be perceived as in a safe state to land and for the runway to be clear to land. Although there have been several technological interventions developed to increase the likelihood that a runway stated as clear to land actually is, the pilots make the final call. Interventions such as runway clear lights and taxiway procedures have decreased collisions related to runway incursions; however, a landing pilot may still fail to detect the warnings. Unfortunately, although the visual system is very good, visual attention failures do occur. In short, runway incursions remain one of the most dangerous events in the aerodrome and, failing all else, is a 100% visual detection task (detailed discussion below).

Early morning in 2010 at the Dunedin International Airport in New Zealand, a patrol officer equipped with flashing roof beacons was conducting his regular parameter scan of the fence around the airport (TAIC, 2010). A portion of the airport service road had partially flooded so he determined he would conduct his scan from the runway. He reported having looked in the direction of any landing aircraft, saw none, and entered the runway with a brief stop to adjust his lights. At this time, 6:08 am, he saw a Metroliner pass approximately 10 meters in front of his vehicle. He reported being unaware of the landing Metroliner until it passed in front of him. Curiously, the Metroliner pilots also failed to detect the vehicle on the runway. The pilots reported detecting the vehicle as they passed by the flashing lights on top of his patrol vehicle. The pilots reported that a collision would have been impossible to avoid if the vehicle had been further onto the runway. A third party, a security agent, reported no adverse visual conditions and saw both vehicles clearly. He reported both the aircraft landing lights and patrol car lights were clearly visible. The agent reported that he did not expect the patrol vehicle to continue onto the active runway. In this example, both parties expected the runway to be unoccupied and failed to detect any information otherwise, such as flashing roof beacons or landing lights. When these people failed to detect the threat in the environment, they each made decisions that could have been deadly if but 10 meters in another direction.

The Federal Aviation Administration lists runway incursions as “an occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designed for the landing and takeoff of aircraft” (FAA, 2020). Historically, fatalities due to runway incursions were common (National Transportation Safety Board, 2007). Safety improvements such as training, improved procedures, runway status lights, and airport surface detection equipment decreased runway incursions and associated fatalities but did not eliminate them altogether. At Logan International in 2009, a departing Airbus A320 narrowly missed striking a construction vehicle that crossed the active runway. Air Traffic Control reported knowing the truck was driving on the taxiway, but the driver unexpectedly failed to follow standard procedure to stop at the runway and request permission to cross. Runway incursion alarms gave the tower no time to react as the truck crossed the runway seconds before the landing aircraft. The driver reported that he thought the runway was closed.

Data show that from 2008-2018, 11,544 runway incursions were reported across 520 airports equipped with an air traffic control tower (FAA, 2020). The FAA reported 1,761 total runway incursions during the 2019 fiscal year. Runway incursions are separated into categories of causation with Operational Incidents, Pilot Deviations, and Vehicle/Pedestrian Deviations (FAA, 2012). A Pilot Deviation incident is a violation of any Federal Aviation Regulation caused by the pilot. The FAA classified 1120 (64%) of the 2019 incursions as Pilot Deviations. The FAA classified 323 (18%) as Operational Incidents wherein below-minimum aircraft separation with other aircraft, vehicles, obstacles, or closed runways was caused by an Air Traffic Controller action. The FAA classified 293 (17%) as Vehicle/Pedestrian Deviation incident wherein pedestrians or vehicles entered any portion of the airport movement areas (runways/taxiways) without air traffic control authorization, and 25 (1%) were reported as Other.

There are four categories of runway incursions ranging from A – D, with Category A being most severe and Category D being least severe (FAA, 2012). Category A is an incident with a narrowly avoided collision; a runway incursion that results in a collision is categorized as an accident. Category B is an incident with significant collision potential marked by a time-critical collision avoidance response. Category C is an incident without time and/or distance pressure for collision avoidance. Category D is a runway incursion incident without immediate threat.

The current experimental scenario uses a runway incursion as the critical stimulus due to the clear and present danger presented to primary task performance and proximity to the anticipated subject fixation point. A Vehicle Deviation Category B runway incursion was selected in the form of a truck crossing the landing runway at the planned touchdown location. As the subject nears the landing runway, a yellow and white truck is visible and stopped on a taxiway that intersects the landing runway. The truck is positioned in a "hold short" orientation, meaning stopped behind the marked threshold at the entrance of a runway. When the subject reaches a specific distance from the runway, the truck begins to move and enters the landing runway. The route the truck proceeds along on the runway intersects with the planned aircraft touchdown location. To increase the detectability of the truck, the truck is a direct threat to landing performance, is in motion, brightly colored, and crosses the natural eye fixation point co-located with the target landing zone.

CHAPTER III

CHARACTERISTICS OF THE INATTENTIONAL BLINDNESS PHENOMENA

A human cannot process the entirety of their sensory experience in exact detail due to limitations in attentional resources (Norman, 1968). Seminal attention studies afforded expansive research areas dedicated to examining the nuances of attention and information processing to understand and predict attentional limitations. For a review article please see Wickens and Carswell, 2012 (suggested original publications include Broadbent, 1977; Kahneman, 1973; Norman, 1968; Posner, 1980; Treisman, 1964; Wickens, 1980). In particular, selective attention enables allocation of these limited cognitive resources to process relevant stimuli and ignore irrelevant stimuli (Broadbent, 1958; Kahneman, 1973; Treisman, 1964). The ability to focus these limited resources to a specific stimulus or set of stimuli allows the human to conserve attentional resources while avoiding over-stimulation. However, this normally helpful process can occasionally filter out information that is relevant to task performance. Failures to detect critical information have been found across senses with visual, auditory, and tactile information examples (Mack & Rock, 1998).

Types of Attentional Failures

There are five major types of visual attentional phenomena that result in the missed detection of visual information, and each have specific parameters that define the circumstances related to the occurrence of that failure. *Inattentional blindness* (IB) occurs when observers fail to notice the presence of a clearly viewable but unexpected event when attentional resources are diverted elsewhere (Hutchinson, 2019; Jensen et al., 2011; Mack & Rock, 1998; Simons & Chabris, 1999). *Change blindness* occurs when an observer fails to detect changes in objects or scenes that occur during a break in visual information such as breaks caused by a saccade, blink, or occlusion (Jensen et al., 2011; Simons & Levin, 1997). *Repetition blindness* is the failure to detect second occurrences of repeated words in visual presentation (Kanwisher & Potter, 1990). *Visual target masking* occurs when one visual stimulus, referred to as the "mask," interferes with the perception of another visual stimulus, referred to as the "target" (Felsten & Wasserman, 1980; Keysers & Perrett, 2002). The *attentional blink* occurs when subjects fail to detect the second of two presented targets when these targets occur within a short time window (180-500ms; Beanland & Pammer, 2011; Shapiro et al., 1997).

The current study focuses on IB. An IB occurrence is typically unexpected and difficult to predict due to the unusual nature of the observation (Mack & Rock, 1998; Simons & Chabris, 1999). The hallmark of an IB occurrence is that observers fail to notice a visual object or event that is clearly visible and easily seen if attention is directed to it. This detection failure is the result of the observer's attention engaged elsewhere and not from aspects of the visual stimulus itself, such as size or visual obstruction. Errors and accidents related to this phenomenon appear across numerous task environments ranging from medical (Lum et al., 2005), aviation (Fischer et al., 1980), nautical (Fraher, 2010), and vehicular (Strayer & Drews, 2007; Strayer et al., 2003).

History of Inattentional Blindness (IB) Research

Mack and Rock (1998) coined the term IB. However, the phenomenon was identified many years prior. Neisser and Becklen (1975) were the first to detect IB when they attempted to produce a visual analog to Cherry's auditory selective attention research. In 1953, Cherry examined the phenomena wherein an individual could be seemingly engaged in conversation but could detect their name spoken in a different conversation across the room. Cherry (1953) conducted a study on the "cocktail party effect" with a dichotic listening paradigm in which subjects were exposed to two streams of auditory information, one in each ear, and told to attend to only one. Cherry sought to identify the level of processing that occurs in an unattended auditory stream. Cherry found that subjects could notice some sematic properties of information contained within the unattended ear, such as their name, and physical characteristics such as the tone. Neisser and Becklen's experiment sought to demonstrate similar findings to that of Cherry for the visual system by developing a selective looking paradigm. Subjects observed a video showing two superimposed semi-transparent videos of people slapping hands or passing a ball. Subjects were instructed to attend only one of these scenes and ignore the other. During the trial, subjects in the unattended scene would commit an unusual behavior, referred to as the critical stimulus. This unusual behavior included handshakes instead of slaps, miming basketball passes, and substituting players with those of the opposite gender. The researchers found that subjects failed to detect this unexpected information contained in the unattended visual channel.

In a follow-up study, Neisser (1979) examined a more obvious critical stimulus. In this study, subjects watched a video of two teams passing a basketball and were instructed to either simply watch the video or count the number of passes among members of a specific group. The critical stimulus was a woman walking through the scene carrying an umbrella (see Figure 1). In this study, 100% of subjects who simply watched the video reported seeing the woman. However, only 48% of subjects directed to count ball passes detected the woman with the umbrella.

Figure 1

Neisser (1979) basketball count task with critical stimulus: woman with umbrella.

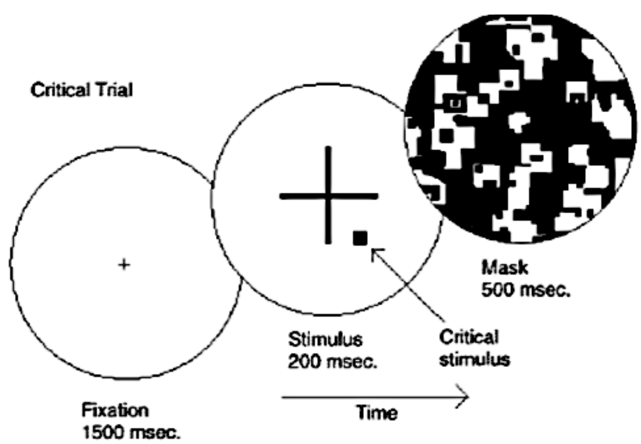
Note. From Neisser 1979 as reproduced in "Gorillas in our midst: sustained inattentional blindness for dynamic events," by D. Simons and C. Chabris, 1999, *Perception*, 28, p. 1063 (DOI:10.1068/p2952). Copyright 1999 by Pion. Reprinted with permission.

In the 1980's, Irvin Rock conducted numerous studies that investigated the relationship of visual perception and attention. Rock found that visual information was not automatically processed in the "preattentive" early stages of visual processing or prior to attention (see Mack & Rock, 1998 for full review). Rock reported that attention must play a larger role in perception to explain subjects unable to detect obvious stimuli or claiming to detect stimuli that were not present (Mack & Rock, 1998). Mack and Rock (1998) argued that the experimental design used to examine preattentive perception, or perception without attention, did not exclude potentially engaging attention. Perception without attention experimental tasks included search protocols such as finding targets within distractors and finding targets while concurrently performing a distraction task. Mack and Rock proposed that the subject is actively seeking the target, so these techniques were assessing divided attention tasks rather than perception without attention. For these reasons, Mack and Rock (1998) developed a new experimental paradigm. In this new technique, the observer did not expect, and was not actively searching for, the critical stimulus. Rather, the critical stimulus occurred while the subject completed another task. Mack and Rock dedicated a decade to the resultant phenomenon.

The main IB experiment featured multiple trials in which subjects viewed a cross shape displayed on a computer monitor and reported which cross arm was longer (see Figure 2). The third trial was the critical trial in which a secondary shape appeared within the viewing area. Here, the researchers asked the subjects to report the longer arm and if any object other than the cross shape appeared during that trial. The fourth trial was a divided attention task trial. Subjects were instructed to attend to both the cross shape and the appearance of an additional object. The fifth trial was a control condition used as a comparison. For this control trial, the subjects were instructed to ignore the cross shape and only attend to the other object (Mack & Rock, 1998).

Figure 3

Mack & Rock's (1998) initial experiment, the cross arm length task with critical stimulus.



Note. From "Inattentional Blindness: An Overview," By A. Mack & I. Rock, 1998, *Psyche, 5*(3), p. 7. Copyright 1998 by MIT. Reprinted with permission.

Like Neisser (1979), Mack and Rock found that during the critical stimulus trial, approximately 25% of subjects failed to detect the unexpected object within their useful field of view, despite later reporting detection in both the divided attention task and control trial. Mack and Rock continued testing IB nuances by varying the critical stimulus features such as location, type, familiarity, and size and found this phenomenon to be robust. In 1998, they published a comprehensive account of this research and labeled the term inattentional blindness (see Mack & Rock, 1998 for review or Mack & Rock, 1999).

Like Mack and Rock, Simons (2000) sought to explore the operational value of perception. Simons argued that perception should be examined with ecological validity and in terms of awareness and performance. He noted that outside of the laboratory setting, observers must be aware of objects in the environment to make volitional changes in behavior. Therefore, in the IB literature, perception is operationally defined as the conscious detection and identification of a stimulus such that the observer is capable of reporting the stimulus (Simons, 2000). The most well-known IB example comes from the Simons and Chabris (1999) "gorilla" experiment (see Figure 3). A modernization of Neisser's work, this experiment did not use semi-transparent or superimposed images but instead showed a single video featuring two teams, identified by shirt color, passing a basketball to members of their own team. Researchers instructed subjects to count the passes made by a specified group. During the video, a woman in a gorilla suit walked into the gameplay area, paused, beat her chest, and then exited. Only 50% of the subjects detected this unusual event.

Figure 4

Simons and Chabris (1999) basketball count task with critical stimulus: gorilla.

Note. From "Gorillas in our midst: sustained inattentional blindness for dynamic events," by D. Simons and C. Chabris, 1999, *Perception*, 28, p. 1063 (DOI:10.1068/p2952). Reprinted with permission.

Complexities in Predicting IB

From the beginning, the IB phenomenon proved to be easy to elicit but difficult to explain with several explanatory hypotheses that either failed to significantly discriminate between detectors and non-detectors or provided counterintuitive results. When discussing IB literature, the most common hypothesis is that subjects failed to detect the critical stimulus because they did not look at it. Overall, researchers found that eye fixation was insufficient to differentiate between detectors and non-detectors. Several studies examined fixation and found that looking does not equate to seeing nor does target proximity to fixation point (Becklen &

Cervone, 1983; Mack & Rock, 1998; Memmert, 2006; Simons & Jenson, 2009; Wood & Simons, 2019).

Becklen and Cervone (1983) examined the impact of fixation on detection in the early days of the attention and perception using a selective looking paradigm. Becklen and Cervone used the paradigm of the woman with an umbrella who crossed directly through the subject's fixation point. The fixation point is the area of a visual stimulus on which the subject focuses. Becklen and Cervone required one subject group to maintain focus on a fixation point and permitted the other to looking freely and found no differences in detection between the groups. Mack and Rock (1998) varied the location of the primary stimulus (the cross) and the critical stimulus. Mack and Rock presented either the cross or the critical stimulus at the fovea (co-located with the centrally located fixation point) or in the parafovea (within 2.3º of the fixation point). In this instance, half the trials featured a critical stimulus that was co-located with the visual fixation point. The fixation point typically corresponds to area of the retina that offers the highest resolution vision (fovea). The parafovea is the area of the retina surrounding the fovea that features reduced visual acuity and increased motion detection. Moving the placement of critical stimulus presentation from parafoveal to foveal resulted in an IB increase from 25% to 85% of the sample, despite the critical stimulus being co-located with visual fixation. Wood and Simons (2019) conducted three studies with moving stimuli and distractors and also concluded that detection was not related to the stimulus crossing the fixation point. They also found that increasing the length of time the stimulus was on screen did not improve detection.

Memmert (2006) extended the work by Simons and Chabris by adding an eye tracker to examine fixation points. To reiterate, Simons and Chabris (1999) asked subjects to count the number of basketball passes among members of a particular team of individuals either wearing black shirts or white shirts. Towards the end of the video, a person in a gorilla suit walked through the scene, paused, thumped her chest, and continued offscreen. Memmert replicated the ball count and detection findings by Simons and Chabris and also found that the time the subject was fixated on the critical stimulus (i.e., the gorilla) was roughly equivalent between those subjects who detected and those who failed to detect the presence of the gorilla. Other researchers also found the amount of time subjects fixated on the critical stimulus was not significantly different between those who detected the critical stimulus and those who did not (Koivisto et al., 2004; Kuhn & Findlay, 2010; Memmert, 2006; Oktay & Cangöz, 2018). Beyond the lack of significant differences in fixation time, the average amount of time a non-detector fixated on the critical stimulus (1 sec) was beyond that sufficient for visual detection. This collection of research demonstrates that spatial attention via eye fixation (looking) simply being co-located with an object or event is not sufficient to assume cognitive awareness (seeing) of that object or event.

A second unsuccessful explanation for IB was that the critical stimulus could be easily missed. Perhaps the critical stimulus was partially hidden, not visible long enough for the observer to detect, or the stimulus moved too quickly. Most et al. (2001) found that 30% of subjects failed to detect drastically different critical stimuli such as a moving red cross in a field of moving black and white targets and distractors (i.e., T and L shapes). The cross was on screen for 5 seconds and differed by color, luminance, shape, and motion trajectory making dissimilarity to other features in the visual field insufficient to assume observer detection.

Another failed hypothesis was related to the length of time the critical stimulus was available for detection; perhaps if the critical stimulus was presented for a longer time the observer would eventually detect. The gorilla of Simons and Chabris (1999) was on screen for 9 seconds. Beanland and Pammer (2011) found a critical stimulus could remain in the attended visual field for greater than 5 seconds without detection. Wood and Simons (2019) conducted three studies that varied the time (1.5, 2.67, or 5 seconds) the stimulus spent onscreen and found no statistical improvement in detection rate as a function of time the stimulus was presented on screen. Wood and Simons concluded that detecting the stimulus typically occurs within the first 1.5 seconds presented or would not occur at all. They found that detection of critical stimuli in sustained IB tasks occurs early in the stimulus onset and is not a slow accumulation of information over time nor evenly distributed across time. Similarly, Simons and Jenson (2009) examined individual differences in the ability to track an object, that is, how fast the subject could track an object at a pre-determined accuracy level, and found that object tracking ability also failed to predict IB susceptibility.

Another unsuccessful explanation for IB was related to the protocol using retrospective self-report to document detection. The standard protocol for IB requires a subject to experience an IB trial and then report if they detected and could identify the stimulus (Mack & Rock, 1999). Neisser (1979) investigated the hypothesis that subjects simply forgot having seen the object or event by the time they were required to report it. Neisser manipulated the amount of time between the critical stimulus occurrence and the detection self-report by adding a 30 second gap after the umbrella woman left the visual field. Neisser found no difference in participants who reported detecting the woman between those who had the 30 second delay and those who had no delay. If a subject saw the woman enough to report it, they could do so despite an additional delay.

Yet another insufficient explanation was that the testing experiments involved simulated or prerecorded events, that an observer would detect the critical stimulus if observed in real life.

In fact, IB instances do occur for events presented in a real-world format, even when those events are highly unusual. Hyman et al. (2010) examined this question with a clown on a unicycle in a university square. Hyman staged a clown on a unicycle in the center of a university campus central square and placed researchers at the edges. The researchers asked students who had passed by the clown and were exiting the square if they had noticed anything unusual during their walk. Individuals just walking or talking with another student reported seeing the clown, but those students who were talking on a cellular phone did not (see Figure 4).

Figure 5

Hymen et al. (2010) path navigation with critical stimulus: unicycling clown.

Note. From "Did You See the Unicycling Clown? Inattentional Blindness while Walking and Talking on a Cell Phone," By I. Hyman et al., 2010, *Applied Cognitive Psychology*, 24, p. 602. (DOI: 10.1002/acp.1638). Copyright 2009 by John Wiley & Sons, Ltd. Reprinted with permission.

This culmination of findings imparts that IB can occur even for objects that intersect the exact location of visual fixation, are dissimilar to other objects in the visual field, are viewable for a long period of time, and are extremely unusual and non-simulated. The physical features of the stimulus are not the cause of IB. The next logical attempt to explain IB is to explore the interaction of the human and the task.

IB in Task Operation

The foundation of the threat of IB is that outside of the laboratory setting, observers must have conscious awareness of objects or events to make volitional changes in planning and behavior (Fischer et al., 1980; Mack & Rock, 1998; Simons, 2000; Simons & Chabris, 1999). IB researchers investigated the concept of perception without attention using two main types of tasks: the basic stimulus response paradigms in visual attention tasks used by Mack and Rock (1998), and the complex naturalistic stimulus tasks like those used by Neisser and Becklen (1975).

IB can easily become deadly when critical objects in the world fail to reach conscious perception. Individuals rely on information received from their interactions with the environment to guide their decision making. An individual experiencing IB has incomplete information with which to generate their mental model. This missing information yields an inaccurate representation of the external world, which limits the individual's ability to assess the accuracy and appropriateness of subsequent decisions. For instance, a pilot may continue to land if they fail to detect a vehicle on the landing runway. In a case such as this, the missed detection of critical stimuli can result in the loss of lives.

In 1980, the National Aeronautics and Space Administration (NASA) examined the impact of a new simulated heads-up-display (HUD) on pilot landing behaviors with a dual crew in a simulated flight study (Fischer et al., 1980). Fischer et al. explored if the HUD would improve landing performance by providing the Pilot Flying (the pilot responsible for landing the aircraft) with an out-the-window view and co-located instrument information. Researchers also examined runway incursion detection with this new system and placed a wide-bodied aircraft on the simulated runway. Regardless of HUD equipage, half of the Pilots Flying failed to detect the single exposure to a wide-bodied aircraft sitting on the landing runway and proceeded to land. This study raised the concern for display-induced attentional shifts in the flight deck (Fischer et al., 1980).

Unfortunately, detection failures that occur in the real world have real consequences including physical damage and loss of life. In 2001, the submarine USS Greenville was to perform a demonstration for distinguished on-board visitors off the coast of Hawaii (COI, 2001; Fraher, 2010; NTSB, 2001). The Captain stated an intention to perform a more exciting demonstration than typical and planned to perform several drastic maneuvers culminating in an emergency ballast blow surface breach. There were several other vessels in the waters and their locations were being charted. Following the high-speed maneuvers, the sonar required a time period to reestablish connection with surface traffic. The Captain of the Greeneville did not wait the defined time period, he incorrectly assumed the last reported coordinates of the known surface traffic were accurate and no vessels presented a threat. He performed a periscope visual scan for surface traffic and reported no visual contacts. Unfortunately, the Captain failed to visually detect the Ehime Maru, a Japanese high school fishery training ship. The submarine surfaced directly under the Ehime Maru causing the vessel to sink, killing nine crewmembers. The Captain performed a visual search scan to identify any hazard for surfacing, paying particular attention to the locations of the expected surface traffic. However, due to numerous mapping inaccuracies and protocol deviations, the expected location for the Ehime Maru was incorrect. Although the Captain was actively searching for surface traffic and he did visually scan the location of the surface traffic, the boat was in a different location than expected and was undetected. Here, IB occurred for a situationally relevant and extremely critical stimulus due to the competition for attentional resources, time pressure, and inaccurate expectation caused by multiple compounding errors.

Simply searching for a piece of information in the visual environment is enough to increase the risk of a detection failure for other threats in the environment. Most and Astur (2007) found expectation as a critical role in IB by demonstrating an increase in IB occurrences for non-target stimuli when operators are actively seeking stimuli. They directed subjects to follow either a blue or yellow navigational arrow found at intersections in a driving simulation (see Figure 5). At the critical intersection, a motorcycle appeared that either matched or did not match the navigational arrow color and turned directly in front of the subject. When the motorcycle color matched the navigational arrow color, only 7% of drivers hit the motorcyclists as compared to 36% when the color did not match. Furthermore, drivers in the ‘‘mismatch’’ condition applied the brakes 186 ms slower on average, and two drivers failed to brake at all. The risk of IB increases when operators are actively seeking other visual stimuli with expected features such as road signage or the expected location of the last known position, like in the case of the Captain of the USS Greenville discussed above and in the case of near collision outside the Moorabbin Airport described below.

Figure 6

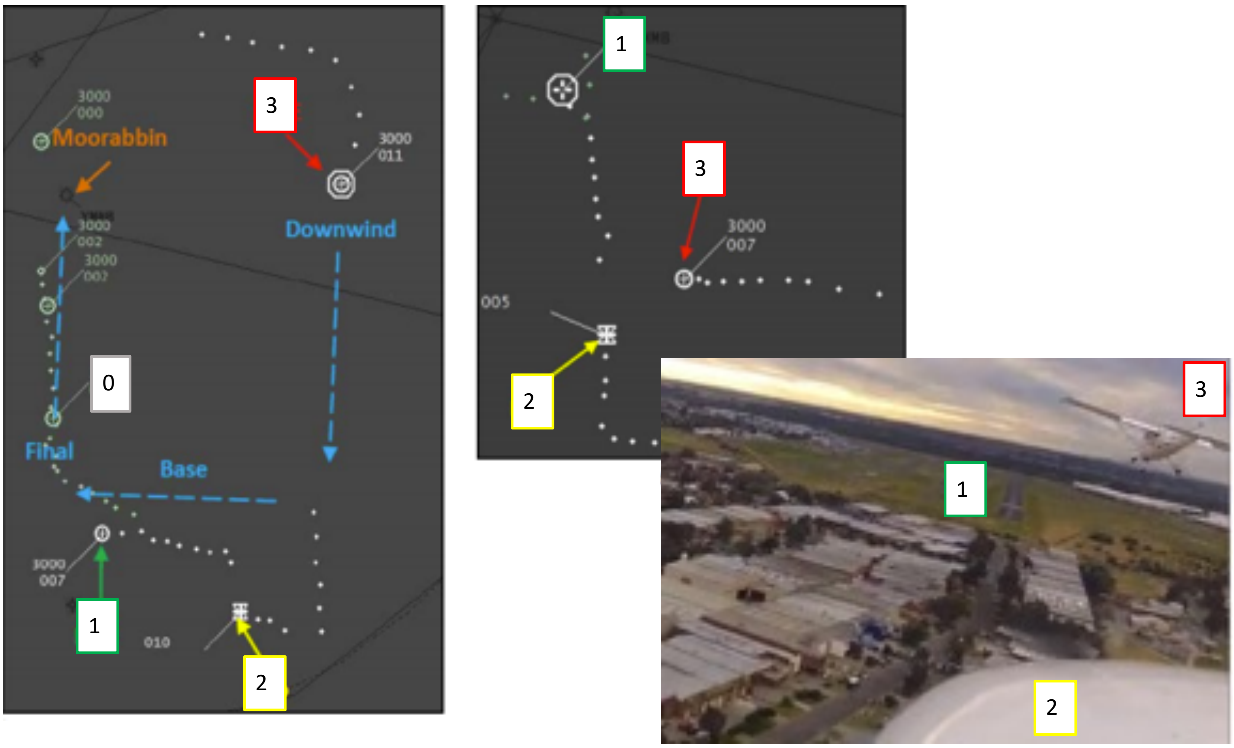
Most and Astur (2007) colored navigational arrows in a driving task.

Note. From “Feature-based attentional set as a cause of traffic accidents,” By S. B. Most & R.S. Astur, 2007, Visual Cognition, 15(2), p. 125-132. (https://doi.org/c9sx27). Copyright 2009 Taylor & Francis. Reprinted with permission.

A pilot and student had a near collision outside of Moorabbin airport in Australia when they mistook one aircraft for another and nearly flew into the missed aircraft. Aircraft 1, 2, and 3 were preparing to land (see Figure 6) (ATSB, 2012). Each were told to visually space behind the preceding aircraft and maintain safe spacing. Aircraft 0 was slow to exit the landing runway. Aircraft 1 had to extend the base leg portion of the flight route to provide enough time for 0 to exit the runway. Aircraft 2 extended the base leg portion of his flight path to maintain a safe spacing from Aircraft 1 causing Aircraft 2 to be much further off the normal flight path than typical when landing on that runway (see Figure 6). Aircraft 3 made visual contact to establish spacing behind Aircraft 2. Unfortunately, the pilot of Aircraft 3 saw Aircraft 1 in the location he expected to be occupied by Aircraft 2 and misidentified Aircraft 1 as Aircraft 2. Aircraft 3 then attempted to space behind Aircraft "2" and nearly collided with the real Aircraft 2 (see Figure 6). Aircraft 3 contained a student pilot and instructor who both reported conducting three separate visual searches for traffic and only gained awareness of Aircraft 2 after the airspace incursion incident occurred. The image to the right in Figure 6 provides a still image taken from the video recorder of Aircraft 2 to show the unsafe proximity of Aircraft 3. The pilots were so close to the actual target aircraft that they not only could see it (as shown by the flight image) but they could have collided with it, and still, failed to detect. This case of mistaken identity highlights the role expectation plays on visual detection and the increased likelihood of IB to an unexpected event.

Figure 7

Near collision outside of Moorabbin Airport.



Note. The image to the left shows the line of aircraft cueing to land with the normal approach path shown in blue. The center image shows aircraft 3 beginning to queue behind aircraft 1 instead of aircraft 2. The image to the right shows a still image taken from the dash recorder of aircraft 2, showing the unsafe proximity of aircraft 3. From “Aircraft proximity event – two Cessna 172S, VH-EWE and VH-EOP,” By Aviation Occurrence Investigation, 2012, Australian Transport Safety Bureau.

Neuroergonomic Approach to IB Assessment

Researchers found that higher attentional demands impair task performance, reduce visual target detection, and increase the likelihood of IB occurrence (Bressan & Pizzighello, 2008; Recarte et al., 2008; Strayer & Drews, 2007) including across modalities, such as auditory (Pizzighello & Bressan, 2008). However, the increased likelihood for IB occurrences were observed at both the low end and the high end of the task demand spectrum. In a two-part study, Simons and Jensen (2009) examined individual differences in task performance ability related to IB. They found that primary task demands influence IB occurrences even when the task difficulty is tailored to produce equivalent task performance for each subject. Bressan and Pizzighello (2008) found that the unexpected event influenced primary task performance but only for those subjects who failed to detect.

As discussed by Driskell et al. (2013), stress and performance are most often studied in terms of overload, with high arousal and high task demand leading to task performance deficits. The current study also identifies the importance of understanding the impact of underload featuring the low task demand and low arousal of boredom. In a simulated driving task conducted by Kennedy and Bliss (2013), subjects who reported higher mental demand while following automated navigational directives were less likely to experience IB to a task relevant critical stimulus than those subjects who reported lower mental demand. However, this task was relatively simple, thus, the high mental demand was within the classification range of moderate workload rather than high, an important difference.

Parasuraman and Rizzo (2008) defined the field of neuroergonomics as the study human brain and behavior in action during work performance. The Dynamic Adaptive Theory (DAT) by Hancock and Warm (1989) defines a steep decline in performance when task demands are very low or very high. The DAT identified an optimal performance comfort zone resulting from moderate stress and instability of performance resulting from hyperstress or hypostress. The DAT model has been further supported in the neuroergonomics literature through the advancements of neurobiology and a better understanding of concentrations of neurotransmitters related to “optimal” executive functioning versus boredom and distress (see Dehais et al., 2020 for an extended review).

During normal flight operation, long periods of inactivity during cruise can create periods of hypostress while other flight periods, such as takeoff and landing, produce a hyperstress environment (Wilson, 2002). The process of landing contains numerous procedures, checklists, and communication interactions, coupled with the increased physical risk related to ground proximity. Simply by nature of the task, the hyperstress related to landing can reduce critical event detection and increase the likelihood of IB. Furthermore, the extended periods of hypostress provide an increased likelihood to succumb to automation complacency and mind wandering.

Humans have a limited capacity for processing information (Broadbent, 1958; Kahneman, 1973). The inability to attend to all things simultaneously has been the focal point of numerous attention and performance models such as Broadbent's Filter Model (1958), Feature Integration Theory (Treisman & Gelade, 1980), Heuristics and Biases (Tversky & Kahneman, 1974), Multiple Resource Theory (Wickens, 2002), and the Malleable Attention Resource Theory (Young & Stanton, 2002). Selective attention is the application of limited cognitive resources. Factors associated with the psychophysiological state of the operator determines how well that attention will be assigned. However, when traditional precursors of degraded task performance, such as effort or workload, are investigated as precursors to IB, myriad counterintuitive and conflicting findings arise. For example, typical predictors such as age, working memory capacity, task workload, primary task performance, and expertise show no significant differences between detectors and non-detectors in one study but significant differences in another (Beanland & Chan 2016; Dehais et al., 2020; Mack & Rock, 1998; Memmert, 2006; Wright, Boot, & Morgan, 2013).

Dehais and colleagues (2020) reported that mental workload represents an interaction between individual differences in capacity and task demands. As described by Dehais et al. (2020), the inconsistent findings in the IB literature may be because IB is not solely a product of cognitive function, but rather, IB appears to sit at the neurophysiological intersection between cognitive function and biological limitation of the visual system, with either having the potential to induce IB. The conflicting findings in the IB literature may actually be related to the task chosen to elicit IB triggering the same response (failure to detect) but caused by different underlying mechanisms. For example, a subject who experiences IB due to a visual system overload has little in common with a pilot landing at a quiet airport who fails to detect a maintenance vehicle sitting on the landing runway with beacons flashing (TAIC, 2010; Wright et al., 2013). And yet, both are examples of the failure to detect visual information that was relevant, detectable, and within the useful field of view.

In this way of thinking, the task used to elicit IB has increased importance. One kind of IB task relies on a densely packed visual field but requires little in the way of higher order cognitive processes (e.g., Wright et al., 2013). This type of task engages the physiological limitations of the visual information processing system to produce the attentional malfunction resulting in IB. Findings from a study that solely used an overloaded visual field to induce IB would generalize poorly to the cognitive complexities and task expectation found when a pilot fails to detect a maintenance vehicle with flashing beacons on the landing runway (TAIC, 2010).

As underlying causes of IB can be different, the metrics required to successfully measure the neurophysiological state that produced that IB will also be different. As previously mentioned in the IB literature review, eye gaze fixation point is typically an appropriate method to determine the location of visual attention in a visual field; except in the case of an IB, in which looking does not equate to seeing. However, information gleaned from eye tracking is still valuable in IB research. Aside from confirming the subject's eyes were open and oriented towards the location at the time the visual stimulus was present, eye tracking can also provide insight into the cognitive state of the operator. Eye tracking offers a way to passively measure eye movement patterns, pupil diameter changes, and eyelid closures which are all used to passively assess human cognitive state without interruption of task performance (Mohan et al., 2019; Peißl et al., 2018).

Pupil dilation is a cognitive state measurement that can be collected passively and in real-time using eye tracking technology (Recarte et al., 2008). Dehais et al. (2020) reported IB to be positively related to fixation duration and negatively related to saccades and fixated areas of interest. Dehais and colleagues (2020) also reported that, although not IB, the related phenomenon of inattentional deafness was negatively related to pupil diameter. Identifying a relationship between eye tracking and IB might permit its use as an automation implementation method or safety feature for operators during task performance in any environment that includes highly visual performance tasks or monitoring, such as aviation.

Wright et al. (2013) found pupil dilation did not predict IB, however, their study may not have had the appropriate design to do so. Wright and colleagues used a common IB induction task to examine how subject effort and primary task engagement impacted IB. Subjects maintained visual fixation at the center of the computer screen and monitored a visual field that contained targets and distractors moving randomly around the field for 60 trials at 8.5s each. At the end of each trail, subjects indicated the number of times a target made contact with the edges of the computer screen. Wright et al. varied workload by number of distractors and by the similarity of the distractors to the target object to be counted in color or shape. Wright and colleagues operationalized effort in terms of pupil dilation and primary task engagement in terms of bounce count error rate. They found a significant negative correlation between pupil dilation and error rate supporting the low/high workload manipulation but found no significant relationship between IB and effort or primary task performance. However, this task contained little in the way of higher-level cognitive processes or decision making. Wright et al. found that reduced pupillary response negatively correlated with higher error rate and higher error rate reflected the higher workload manipulation, however, neither demonstrated a significant difference between the detectors and non-detectors. Wright et al. explained these findings by reporting the effort recruited for the primary task performance was unrelated to effort in noticing expected events.

Wright et al.'s results might be more readily explained considering the neurophysiological conceptualization of IB. Wright et al. manipulated task workload by capitalizing on increased visual load, but the task was monotonous and repetitive, the protocol did not permit free eye movement, and no subjective measures of workload or effort were administered for comparison. Kahneman and Peavler (1969) found that the motivation of the subject impacts the magnitude of pupillary response. Franklin and colleagues (2013) reported that the mind will wander in the absence of a goal-directed task and that pupil dilation will also occur in response to the mind wandering content, rather than content of the task and stimulus presentation. Beatty (1982) noted that a critical role of intentional attentional engagement was related to effort (reviewed in Beatty & Lucero-Wagoner, 2000). Ultimately, an unengaging task may lead to inconclusive pupil dilation results. In a task by Kang et al. (2009) found that subjects curious about the answer to a trivia question answer produced a small but detectable (8% vs. 4%) pupil dilation response (Kang et al., 2009).

Mathôt (2018) indicated that pupil dilation is indeed reflective of cognitive state and found that a small-sized pupil was related to drowsiness, a moderate-sized pupil related to focused attention, and a large pupil characterized increased cognitive activation. Winn et al. (2018) identified that task-evoked pupil dilation was not a single simplistic construct summed up in effort, but instead is reflective of the intersection of many internal processes such as attention, engagement, arousal, anxiety, and effort. Therefore, the potential for determining the use of pupillometry in predicting IB may be better examined using an engaging flight paradigm with a naturalistic and unobtrusive collection method.

Recarte and Nunes (2000, 2003) conducted a real-world driving study that included a secondary task and found pupil dilation varied with increased secondary task workload. Recarte and Nunes (2003) replicated the pupil dilation findings and correlated them with the subjective workload measure to support that pupil dilations were consistent with subjects' perceptions of the effort applied to the task. In a follow-on, Recarte et al. (2008) reported that during complex task performance such as driving, the pupil dilation represents the entire mental workload experienced which included the workload associated with the current task operation and the workload associated with planning to complete upcoming tasks (e.g., passing another car, way-finding).

Using neurophysiological conceptualizations, the use of pupil dilation as an indicator of IB-predisposing cognitive state conditions would be more appropriately assessed when coupled with a cognitively complex IB task. Meaning, an experiment that uses pupil response as a method to examine cognitive state conditions surrounding IB occurrences during task operation should use an experimental task that engages similar neurophysiological features as an operational environment, such as an aviation simulation.

Purpose of Current Study

Data for this study were obtained as part of a larger study examining the use of a specific type of EEG analysis to detect IB. These data came from the first IB flight simulation run and prior to any other experimental exposure. The other variables in this study included aircraft performance, eye tracking, subjective report questionnaires, NASA-TLX Workload Rating Scale, and overall operator state via psychophysiological assessment. Eye tracking, NASA-TLX, and subjective report questionnaires are reported herein.

The current study examines the cognitive state conditions associated with an IB occurrence to an operationally relevant unexpected object. The goal of this study was to increase understanding of the predictors related to visual stimulus detection during complex task performance to provide increased insight into pilot readiness to respond to an unexpected event. Specifically, this study examined IB occurrences to an unexpected runway incursion during a simulated flight landing task across three automation conditions with varied task load (full automation, partial automation, and manual) to examine pupil dilation as a discriminator between detectors and non-detectors. This study examined the use of eye tracking to identify predictors of unexpected visual stimulus detection during simulated flight task performance to explore the potential for real-time IB mitigation opportunities.

The following hypotheses correspond to the preceding review of the IB literature:

- *H1.* Subjective task demand scores on the NASA-Task Load Index (TLX) would vary significantly as function of automation condition. Specifically, subjects in the full automation condition would report the lowest overall task demand, subjects in the partially automated condition would report higher task demand than the automated condition but lower task demand than the manual condition, and subjects in the manual condition would report the highest overall task demand (Driskell, Driskell, & Salas, 2013; Hancock & Warm, 1989; Hart & Staveland, 1988; Recarte et al., 2008).
- *H2.* The likelihood of IB occurrence would vary across automation condition such that the partial automation condition would have the lowest likelihood of IB occurrence as compared to the full automation and manual condition. The manual condition was expected to have higher likelihood of IB occurrence than partially automated condition due to higher cognitive load (Cartwright-Finch & Lavie, 2007). The full automation condition was expected to have higher likelihood of IB occurrence than the partially automated condition due to automation complacency and mind wandering (Bailey & Scerbo, 2007; Franklin et al., 2013; Parasuraman & Manzey, 2010).
- *H3.* Subjects who indicated higher task demand (i.e., reported higher scores on the overall NASA-TLX) would exhibit greater pupil dilation as compared to subjects who scored the task as lower in demand (i.e., reported lower scores;). Task demand will be positively associated with pupil dilation, such that participants who report higher scores on the overall NASA-TLX will also exhibit greater pupil dilation (Recarte et al., 2008).
- *H4.* Subjects with larger pupil dilation would have decreased likelihood of IB occurrence compared to those with smaller pupil dilation (Beatty & Lucero-Wagoner, 2000; Dehais et

al., 2020). Specifically, those subjects with larger pupil dilation were more likely to detect the critical stimulus than those subjects with smaller pupil dilation.

CHAPTER IV

METHOD

Subjects

Subjects were 60 non-commercial aviation pilots (29 male, 31 female) with a mean age of 34.5 years (*SD* = 13.3 years, range = 20 to 64). Non-pilots were chosen for this initial study as the required sample size was cost prohibitive for a commerical pilot sample and to provide support for a future study with a certified pilot sample. A power analysis (as described in detail in the Data Analysis section below) indicated a required sample size of 22. Subjects were required to have normal or corrected-to-normal vision and hearing assessed via self-report on the demographic questionnaire. Most participants self-reported as right hand dominant (n=56), 2 participants self-reported as left hand dominant, and 2 as ambidextrous. Mean computer usage per day was 5.5 hours (*SD* = 3.2 hours; range = 0 to 15), PC game use was reported as 6.9 days per month (*SD* = 9.7 days; range = 0 to 30). Twelve subjects reported playing simulated flight games during PC game usage at a mean of 7% of their total estimated gaming time (*SD* = 4.8%; range = 1% to 20%).

As previously noted, these data were obtained as part of a larger study examining the use of EEG to detect IB. Due to the strobing lights used for the EEG photic stimulation portion of the full experiment, the participants were required to be over the age of 20 with no history of epilepsy or recent traumatic brain injury. The strobe frequencies were within the trigger range of photosensitive epilepsy and the age of onset is typically before the age of 20 (De Bittencourt, 2004).

Non-civil servants (n =59) were compensated with $50. Civil servants (1.66%, n = 1) participated in the research in their official capacity as Federal employees and were not compensated as per NASA policy. Given that only one civil servant participated in this study, comparisons between the civil servants/non-civil servants were not possible. Subjects were recruited using the NASA Langley contractor tasked with human subject recruitment for studies. The recruitment was made by posting the opportunity on an online recruitment database available to the public (https://flight-research.larc.nasa.gov/) as well as a physical flyer posted in several common spaces at the research facility (see Appendix H). The participant recruiter matched participants to available schedule slots, obtained visitor badges for the participants to enter the research center, escorted the participants from the badge office to the research laboratory, and handled all aspects related to the participant payments.

Prior to data collection, this study was approved by the Institutional Review Board at NASA Langley Research Center and two internal branch experiment reviews (preliminary experiment review, final experiment review). Both branch review formats included a presentation to a branch panel with an open audience and question and answer period. Completion of both branch reviews required approval of the team lead, the branch panel, and a member of branch management. Prior to participation, subjects read and signed an informed consent document (see Appendix A) and a Privacy Act Notice (see Appendix I) and were provided a copy of each to keep. Participation was entirely voluntary; participants gave consent with the understanding that they could end the study at any point.

Brief Summary of Experimental Design

Subjects were randomly assigned to one of three automation conditions. These automation conditions required subjects to monitor or operate the aircraft thrust and attitude controls as conceptually similar to autopilot (fully automated), auto-throttle (partially automated), and manual (manual) flight conditions. After three training sessions, subjects then completed the final five minutes of a simplified landing scenario. The task was a simplified approach to an airport and included two holding speeds, one altitude-specific speed change, and orienting the aircraft to land on a specific runway. At approximately 10 seconds prior to touchdown, a truck (the critical stimulus) moved along a taxiway and onto the landing runway at the touchdown target location. The scenario was suspended, and the simulation displays were blanked just prior to touchdown (i.e., prior to collision with the vehicle). Subjects then completed the post-experiment IB questionnaire and NASA TLX.

Materials

Subjects completed a background questionnaire that included relevant demographic information including age, sex, information about vision and hearing, and flight simulator experience (see Appendix B). Researchers then provided a description of the study, the flight simulator, the scenario, the automation condition, and experimental instructions.

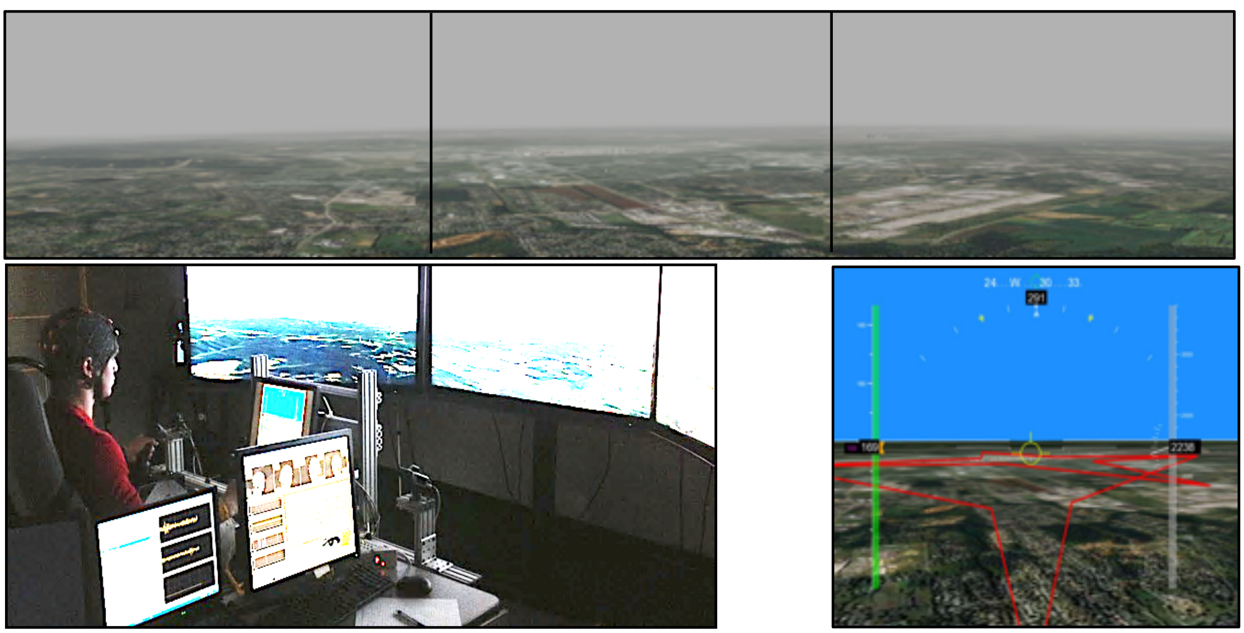
Experimental Manipulation. The experimental manipulation for this study was flight control automation. Subjects were randomly assigned to one of three automation conditions and remained in that automation condition for the entire experiment. The order of random assignment was pre-defined using a random sequence number generator prior to experiment launch to ensure equivalent category membership. Three automation conditions were chosen for the conceptual similarity to autopilot, auto-throttle, and full manual operation. The conditions were full automation, partial automation, and manual (i.e., no automation). Subjects in the full automation condition were instructed to monitor the automation-controlled flight path and speed and to report any deviations from expected flight parameters. Subjects in the partial automation condition manipulated the attitude of the aircraft and monitored the auto-throttle control of the speed changes. Subjects in the manual condition manipulated both the aircraft attitude and thrust.

Subjects performed three practice sessions within their assigned condition to achieve task proficiency. Training is discussed further in the procedure section below.

System Description. The flight simulator used for this study was capable of varied levels of fidelity and flight control accuracy. This experiment utilized a simplified set of flight controls to accommodate non-pilot subjects. This simulator included force feedback sticks to enable both visual and tactile feedback related to the state of the aircraft control settings including those changes made by automation. The simulator visual environment provided an out-the-window view and a simplified primary flight display (see Figure 7).

Figure 8

Experimental set-up: out-the-window and primary flight display during speed change.



The out-the-window view consisted of three 60 Hz 65" Sharp LCD monitors placed in an arc 8 ft. from the subject with the two side screens turned towards the subject to provide equidistance viewing. The simulated environment featured a moderate-to-high definition resolution rendering of the Louisville International Airport and surrounding area. The primary flight display was placed on one 60 Hz 17" Dell monitor at a fixed distance of 26" from the subject. The primary flight display contained a simulated synthetic vision rendering of the out-the-window image and a basic aircraft instrument package including flight path marker, speed, altitude, and heading information. The primary flight display included a virtual waypoint overlay represented by a three-dimensional red star located at the route coordinate at the target altitude of the required speed change (see Figure 7).

The Task. The simulated aircraft flight dynamics model was a twin turbo-prop commuter plane Dash-8. The flight scenario consisted of daytime flight conditions, overcast, with 3 miles or greater of visibility. Turbulence was represented by a pre-recorded light wind created via a randomly generated seed and a sum of signs algorithm such that all subjects experienced the exact same conditions without a discernable pattern.

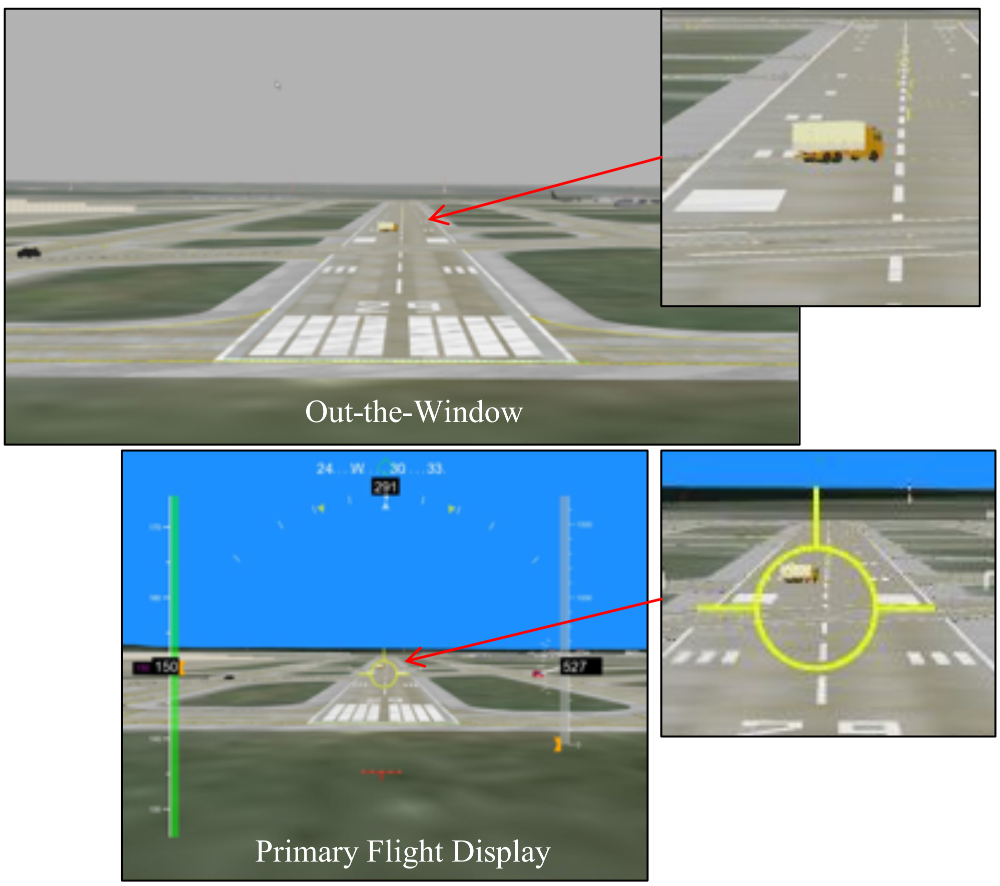
The flight task required subjects to experience the final five minutes of a simulated simplified landing scenario by either piloting or monitoring an aircraft to land on runway 29 at Louisville International Airport. The scenario initialized during mid-approach at approximately 5 minutes before touchdown. The total run from starting point to touchdown point covered a Euclidean distance of 48,925 ft. Perfect performance of this scenario required continual descent from 4890 ft to 462 ft onto runway 29. The specified airspeed was 180 knots true airspeed (KTAS) until 2200 ft, then reduce speed to 150 knots true airspeed until touchdown. The simulation ended and the simulation displays were cleared prior to actual touchdown.

Critical Stimulus. A runway incursion is defined as any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designed for the landing and take-off of aircraft (FAA, 2012). There were seven vehicles in the proximity of the landing runway: three nonmoving, three moving, and one critical stimulus. These vehicles were in view for approximately 40 seconds and only the critical stimulus provided a conflict to landing. The three moving vehicles were on the two taxiways parallel to the active runway. Vehicular motion was activated when the subject reached a pre-specified position in the scenario rather than based on time elapsed. These vehicles triggered at a Euclidean distance of 6510 ft from the target touchdown point and were in-motion for approximately 26 seconds.

The critical stimulus took the form of a Vehicle Deviation runway incursion of an orange and white box truck crossing the landing runway at the touchdown target point. The critical stimulus was a Category B runway incursion because a trained pilot could avoid a collision by performing a go-around maneuver, once detected. The current study examined the detection of a stimulus and not the response actions to avoid the stimulus; meaning, this goal of this study related to whether or not the subject detected the vehicle on the runway, and not if the subject performed the right action following detection. Therefore, the non-pilot subjects were not expected to perform any avoidance maneuvers, were not trained to do so, and the simulation ended prior to the subject colliding with the vehicle (Mack & Rock, 1998). Any spontaneous avoidance maneuvers were simply recorded. The truck was positioned on an intersecting taxiway at a hold-short position immediately perpendicular to the landing runway. When the subject reached 2590.8 ft (Euclidean distance) from the touchdown point, the truck triggered into motion, entered the active landing runway and presented a direct collision threat to the landing aircraft (see Figure 8). The path of the truck was designed to pass through the natural fixation point and simulate a parafoveal then foveal then parafoveal presentation. The truck was in motion until the end of the run which was approximately 10 seconds. The scenario ended and displays were cleared just prior to touchdown and the subject completed the post-experiment IB questionnaire.

Figure 10

Out-the-window and primary flight display with Critical Stimulus.



Note. Participant view of the critical stimulus in each flight display.

IB Questionnaire. Following the scenario, subjects completed via pencil and paper a hardcopy self-report questionnaire designed to elicit responses related to the scenario (see Appendix C). The established assessment of IB is the failure of a subject to consciously perceive the critical stimulus such that they are unable to report detection of the stimulus. Consistent with Mack and Rock's (1998) IB paradigm, the post-experimental self-report questionnaire specifically prompted the subject to report detection of the critical stimulus. This technique is reproducible, supported, and accepted as the experimental paradigm for IB detection and assessment (Mack & Rock, 1998; Memmert, 2006; Most et al., 2005; Neisser & Becklen, 1975; Simons & Chabris, 1999; Varakin et al., 2004).

This questionnaire provided these questions to assess IB: "*Did you see anything on or above the landing runway? Yes or No*" and *"If so, please describe"* for the subject to report detection. Subjects were considered as unable to detect the critical stimulus and classified as exhibiting IB if they indicated either a "no" to the post-experimental questionnaire or a "yes" but could not accurately describe the truck (Mack & Rock, 1998; Most & Astur, 2007). The dependent variable, detection or non-detection of the runway incursion, was measured dichotomously: 0 = detected (non-IB case), 1 = non-detection (IB case). To reduce data entry error, the completed questionnaires were digitized using the double entry method. The researcher and a laboratory assistant both independently entered these data into separate Excel sheets. These sheets were compared, and any data discrepancies were resolved through examination of the original completed questionnaire.

To avoid conceptual cueing, the questionnaire contained other questions related to the preceding scenario such as speed/altitude changes, landing intentions, and visual scenery items. These non-IB fill-in-the-blank questions were examined for any subject statements that indicated detection of the critical stimulus to ensure accurate IB/Non-IB categorization. The previously presented results of these data a response was categorized as detect if the subject accurately described the critical stimulus on the two specific IB questions listed above. In this study, a response was categorized as detect (non-IB; coded as 0) if the subject accurately described the critical stimulus in any answer space beyond just the two targeted IB items, and a non-detect (IB; coded as 1) if there was no indication on the questionnaire that they saw the stimulus.

NASA-Task Load Index (NASA-TLX). Hart and Staveland (1988) defined workload as the cognitive resources required for an individual to perform a task at a specific level. Subjects completed the NASA-TLX (see Appendix D) to provide a subjective rating of perceived workload. The NASA-TLX is a multi-dimensional scale of workload with six subscales that assess a different dimension of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration level (Hart, 2006; Hart & Straveland, 1988). Recarte et al. (2008) identified the NASA-TLX as the best predictor of visual detection impairment when compared to eye blink and pupil dilation in dual-task conditions of visual and mental workload. Each subscale has a single item scored on a scale of 0-20 (see Appendix D). The NASA TLX has a test-retest reliability of r = .83 (Hart & Straveland, 1988) and a Cronbach's α of 0.84 demonstrating high internal consistency (Flägel et al., 2019). Rubio et al. (2004) reported high convergent and concurrent validity in mental workload, task performance, and predictive task performance with the TLX correlating above 0.97 to two other accepted measures of subjective workload: Subjective Workload Assessment Technique (SWAT) and the Workload Profile (WP) (Rubio et al., 2004).

After completing the simulated flight task, subjects indicated the individual magnitude of the six workload elements by marking a position along a corresponding line with 20 equally spaced hash marks. For this study, the NASA-TLX had a high level of internal consistency and found the same Cronbach's alpha of 0.84 as reported by Flagel et al. (2019). The procedure outlined by Hart (2006) was used to produce the total TLX score by averaging the six subscale scores without the additional weighting scale of the original TLX.

Smart Eye™ Eye Tracking System. The Smart Eye Pro™ eye tracking system, version 5.0, was installed in this simulator and used to capture eye behavior information, pupil dilation in particular. For detailed information and description of the eye tracker, eye tracker model, and eye tracking procedure please see Appendix J. The Smart Eye (SE) is a multi-camera head and gaze tracking system that enables naturalistic data collection and uses features of the head and face to calculate the head pose and gaze direction of subjects in a defined 3D space relative to the position of the tracking camera (Ahlström, Kircher, & Kircher, 2009; Ellis, 2009; Smart Eye, 2008). The experimental system consisted of four 60 Hz Sony HR-50 6.0mm lens cameras connected by ethernet, two IR-diode flashers, a 3-dimensional virtual representation of the critical portions of the testing environment (world model), and a 3-dimensional virtual representation of the subject’s head (head model). The SE uses infrared diodes to provide consistent lighting across the subject’s face and to produce glints, or cornea reflections of the IR flashers, that the system uses to locate the eye center. Multiple studies demonstrated successful pupillometry conducted in real-world driving applications with normally occurring lighting variations and free head movement (Recarte & Nunes, 2000, 2003). In this study, the subject was able to move freely, therefore, an accurate head model across the expected gaze range of the environment was required for accurate pupil diameter measurements.

Despite the best configuration, specific instances can preclude successful eye tracking such as extended eyelid closures, extreme head angles in any direction, exiting the eye tracking location, characteristics of the iris/pupil, and characteristics of some glasses and contacts. An acceptable eye tracking profile was created for all subjects in this study. The pupil dilation data cleaning procedure is described in the results section below.

Procedure

In the original study, all subjects reviewed and signed the IRB-approved Informed Consent form and completed the background questionnaire. Subjects were provided a short experimental overview briefing regarding the flight task, the automation condition variations, and any physiological recording equipment used for workload response collection. Subjects were equipped with the physiological monitoring equipment and an eye tracking profile was created. Subjects were randomly assigned to one of three automation conditions. Subjects were fitted with physiological sensors, seated, and positioned in the simulator and began the Smart Eye head model protocol.

To ensure adequate task performance competency with the equipment in the environment, after being randomized to one of the three automation conditions, the subject completed a simulator familiarization procedure and three training runs. The subject maintained the same automation condition throughout training and testing. The training sessions utilized the same experimental environment with a Northerly approach and landing on Louisville International Airport runway 35L with similar speed and altitude changes. The training runs consisted of three flights to landing, two flights with researcher assistance as needed and at least one solo final run. The guided training sessions permitted the researcher to provide verbal or physical guidance as needed to assist the subject while learning task operation. The final training run was completed entirely by the subject to demonstrate task proficiency. All subjects completed training to proficiency in three runs. All training scenarios ended just prior to imminent touchdown. This was because the test run included a vehicle on the runway that would result in a runway collision if the subject was permitted to land. A runway collision would make the critical stimulus overtly obvious and likely eliminate any odds of a subject failing to detect. The subjects were told the scenario would end just prior to landing due to limitations of the virtual environment. No vehicles were on or near the runway during the training sessions. The eye tracking model was validated, and any necessary modifications were made.

Next, subjects completed the experimental scenario and guided the aircraft down to Louisville International Airport runway 29 using a Westerly approach with the control/monitoring combination for the automation condition as trained. The scenario ended immediately prior to touchdown to avoid potential collision between the aircraft and the critical stimulus. Following the end of the the scenario, the subjects completed the post-experiment IB questionnaire and the NASA-TLX form. This portion of the testing session pertinent to the current study lasted approximately one hour. Subjects completed the remainder of the full experiment and were debriefed.

CHAPTER V

RESULTS

Power Analysis

All power calculations were made using G*Power (Faul et al., 2007). The power analysis for the ANOVA was conducted using G*Power. Recarte et al. (2008) identified the overall NASA-TLX as a retrospective predictor of visual detection impairment in dual-task conditions of visual demand with an $\eta^2 = .566$. For the ANOVA, G*Power ANOVA: Fixed effects, omnibus, one-way, *F*-test was conducted with a calculated effect size f of 1.14993, alpha of .05, power of .80, and number of groups of 3 to yield a required total sample size of 12.

For the logistic regression analysis, G*Power logical regression z-test was conducted for two-tailed, odds ratio = 11, H0 proportion of .45, alpha of .05, power of .80, normal distribution, yielding a total required sample size of 22. IB literature was used to determine the required estimation of the predicted odds ratio and the smallest proportion of IB proportion of cases. The Cartwright-Finch and Lavie study (2007) was selected due to the similarity of cognitive load manipulation and IB. Cartwright-Finch and Lavie found 90% (18 of 20) of subjects experienced IB in the high cognitive load condition as compared to 45% (9 of 20) of subjects in the low cognitive load condition. This information is organized in the following diagram:

	Case – IB	Non-Case – Detect
High Workload	18	2
Low Workload	9	11

The odds ratio formula (a*d)/(b*c) was populated using Cartwright-Finch and Lavie (2007) data from the table above as (18*11)/(9*2) and yielded an odds ratio of 11. This group difference is similar to that found by IB researchers (Simons & Chabris, 1999, Simons, & Jensen, 2009).

Pupil Dilation Data Cleaning

All statistics and analyses were conducted using IBM SPSS Statistics version 25. The current study included four main hypotheses. Significance was determined by a *p*-value less than .05 level. The eye tracking metrics examined the eight seconds following critical stimulus activation (i.e., the truck starts to move) to ensure capture of the pupillary response after detection. Pupillary response ranges from 500ms to 1.5s and emerges, on average after 1s (Winn et al., 2018). As discussed in the introduction, subjects who detected an IB stimulus were likely to do so within the first 1.5 seconds (Wood & Simons, 2019).

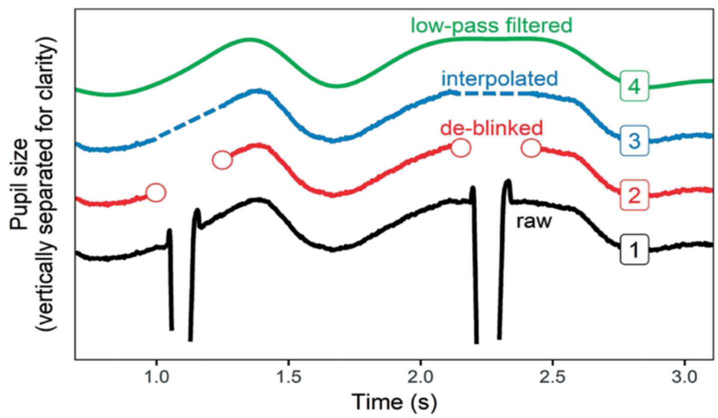
The critical stimulus was triggered to begin moving based on the distance of the subject to the runway rather than simulation time to ensure all participants experienced a similar threat to landing. Therefore, the simulation time for the critical stimulus onset was slightly different for each participant. The data were aligned using the critical stimulus onset as a shared zero point. The examination window included the zero point and the following 480 data points (approximately 8 seconds). This range was considered an acceptable time window to detect a possible response based on the time documented by Winn et al. (2018).

The pupil diameter for the left and right eye was recorded in meters and was converted to millimeters (mm) for consistency in reporting. Next, the data from the left and right eye of each subject were each assessed for accuracy and completeness. This study design featured an applied setting with an interest in subject pupillary response during performance of a simulated aircraft landing scenario. To this end, subjects were permitted to look and move freely with the expectation that portions of subject data would likely require exclusion during the data cleaning. Accuracy was assessed using the expected range for pupil dilation and the Smart Eye data quality metric. The human pupil diameter ranges from approximately 2mm to 8mm (Mathôt, 2018). Pupil diameter measurements outside of 2 to 8 mm were considered outside of typical physiological range due to measurement error and excluded from analysis. For data quality, Smart Eye provides data from the pupils of each eye along with a quality metric that is normalized from a scale of 0.0 to 1.0 (Smart Eye 2008). Following the protocol by Ahlström et al. (2009), data with a quality rating lower than 0.25 were excluded from analysis.

Next, the left and right eye data that were both a) within the defined pupil dilation range and b) had a data quality score greater than 0.25, were averaged to produce a single mean value. Missing data were expected due to eye blinks or measurement error. Using BioPack Student Lab v4.1, interpolation was used to replace missing data followed by a 10 Hz low pass filter as documented in the procedure for pupil dilation data cleaning documented by Winn et al. (2018; see Figure 9).

Figure 11

Winn et. al., (2018) sequential steps of data processing for pupillometry data.



Note. 1) Raw pupil dilation data (black) with missing data, 2) remove transient excursions (red), 3) interpolate gaps 4) lowpass filter data (green). From "Best Practices and Advice for Using Pupillometry to Measure Listening Effort" Winn et. al., 2018, *Trends in Hearing, 22*(1). https://doi.org/10.1177/2331216518800869. Permission not required for reprint.

The data cleaning process for pupil dilation reduced the sample size for Hypothesis 3 and Hypothesis 4 from 60 participants to 50 participants. Data from 10 participants were excluded for the following reasons: data from 8 participants were excluded because more than 20% of their data were missing or categorized as measurement error (Winn et al, 2018), data from 2 participants were excluded due to missing either the eye tracking data file or the critical stimulus onset information required for data alignment.

Data Analysis

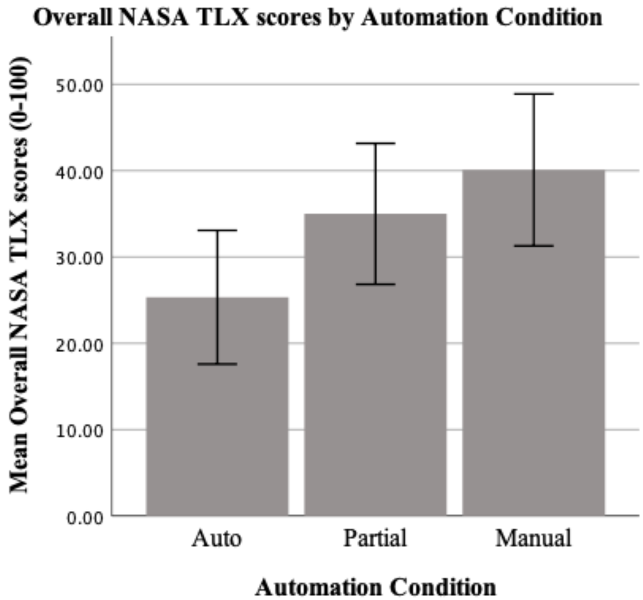
Hypothesis H1. Subjective task demand scores on the NASA-Task Load Index (TLX) would vary significantly as function of automation condition. Specifically, subjects in the full automation condition would report the lowest overall task demand, subjects in the partially automated condition would report higher task demand than the automated condition but lower task demand than the manual condition, and subjects in the manual condition would report the highest overall task demand (Driskell, Driskell, & Salas, 2013; Hancock & Warm, 1989; Hart & Staveland, 1988; Recarte et al., 2008).

A one-way ANOVA was conducted to determine if the subjective overall task load (overall TLX scores) was different across automation conditions (full automation, partial automation, manual). ANOVA has six assumptions. Three of these assumptions are related to the study design: one continuous dependent variable, one categorical independent variable with two or more independent groups, and independence of observations. The other three assumptions are testable with statistics: no significant outliers, approximately normal distribution, and homogeneity of variance. Outliers were assessed for the overall NASA TLX scores using boxplots. An observation greater than 1.5 interquartile ranges from the edge of the box was considered an outlier. No outliers were detected. Due to there being fewer than 50 subjects per condition, Shapiro-Wilk test for normality was used and found normal distribution ($p > .05$). Homogeneity of variance was confirmed, as assessed by Levene's test for equality of variances ($p = .803$).

The subjective overall task load (TLX score) was lower in the full automation ($M = 25.33$, $SD = 16.55$), then higher in the partial automation ($M = 35.0$, $SD = 17.44$), to highest in the manual ($M = 40.08$, $SD = 18.81$) automation condition groups (see Figure 10). Table 1 provides the mean, standard deviation, confidence interval, and minimum/maximum observed scores for the overall NASA TLX means across automation conditions. The overall NASA TLX score was statistically significantly different for different levels of automation condition, $F(2, 57) = 3.62$, $p = .033$, $\eta^2 = .113$.

Figure 12

Overall NASA TLX scores by automation condition: full auto, partial auto, manual.



Note. N =20 per group, total = 60. 95% CI Error Bars.

Tukey post hoc analysis revealed the difference in overall NASA TLX score from full automation to manual (14.75, 95% CI [1.34, 28.16]) was statistically significant (p = .028). The difference in overall NASA TLX mean scores from the full automation to partial (9.7, 95% CI [-3.74, 23.08]) conditions was not significant (p = .201) and from partial to manual (-5.1, 95% CI [-19.07, 8.91]) conditions was also not significant (p = .652).

Table 1

Descriptive statistics for the Overall NASA TLX scores by Automation Condition.

Condition	*M*	*SD*	95% CI for Mean		Min	Max
			LL	*UL*		
Auto	25.33	16.55	17.59	33.08	0	56.67
Partial	35.00	17.44	26.84	43.16	5.83	59.17
Manual	40.08	18.81	31.28	48.88	10.83	79.17
Total	33.47	18.39	28.72	38.22	0	79.17

Note. *N* =20 per group, total = 60. CI = confidence interval; LL = lower limit; UL = upper limit.

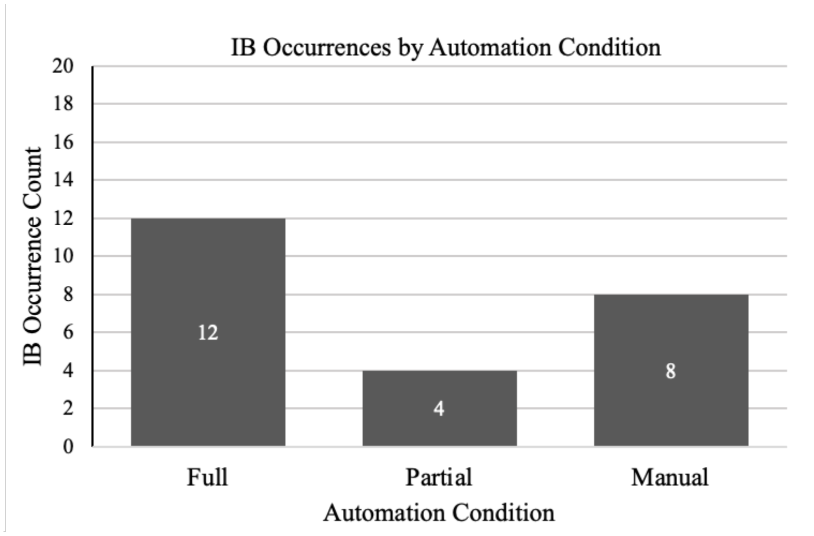
Hypothesis H2. The second hypothesis was that the likelihood of IB detection would vary across automation condition, with the partial automation condition exhibiting the lowest likelihood of IB occurrence as compared to the full automation and manual condition. The manual condition was expected to have higher likelihood of IB occurrence than partially automated condition due to higher cognitive load (Cartwright-Finch & Lavie, 2007). The full automation condition was expected to have higher likelihood of IB occurrence than the partially automated condition due to automation complacency and mind wandering (Bailey & Scerbo, 2007; Franklin et al., 2013; Parasuraman & Manzey, 2010).

A chi-square test was conducted between automation condition (full automation, partial, manual) and IB occurrences (detect or fail to detect). There are five assumptions associated with a chi-square test. The first four are related to study design and were met: one dependent dichotomous variable, one independent categorical variable with three or more categories, independence of observations, and a single observation using random assignment. The fifth assumption is that each expected cell has greater than 5 observations. This assumption was met as determined by a 2 x 3 crosstabulation providing a minimum expected cell count greater than five.

Ultimately, 12 (60%) subjects in the full automation condition failed to detect the critical stimulus compared to 4 (20%) subjects in the partial automation condition and 8 (40%) subjects in the manual condition, which is a statistically significant difference in the proportion of IB, $p = .036$. To determine which of the three conditions significantly differ, the post hoc analysis z-test of two proportions was conducted using pairwise comparisons. The proportion of subjects who failed to detect the critical stimulus in the full automation condition was statistically significantly higher than the partial automation condition, $p < .05$. The proportion of subjects who failed to detect in the manual condition was not statistically significantly different from the full automation or partial automation, $p > .05$ (See Figure 11).

Figure 13

Detect (no IB) vs Non-Detect (IB) by Automation Condition.



Note. N =20 per group, total = 60. Count in table is number of subjects who experienced IB in that automation condition.

Hypothesis H3. Subjects who reported higher task demand (i.e., reported higher scores on the overall NASA TLX) were expected to have increased pupil dilation as compared to subjects who reported low task demand (Recarte et al., 2008). A linear regression was conducted between the NASA TLX and the pupil dilation scores.

Prior to data analysis, the seven assumptions of linear regression were evaluated. The first two assumptions are that the independent (predictor) and the dependent (criterion) variables were measured continuously. The next five assumptions were examined with statistical analysis: a linear relationship between independent and dependent variables, independence of observations, no significant outliers, homoscedasticity, and normal distribution of residuals.

Linearity between variables was examined using scatterplot of overall NASA TLX (X-axis) against average pupil diameter in mm. Visual inspection of this scatterplot indicated a linear relationship between the variables. Independence of observations was expected due to experimental design and independence of residuals was confirmed by a Durbin-Watson statistic of 1.82. No obvious outliers were identified in the scatterplot and none with standardized residuals ±3 were detected in the casewise diagnostics. Homoscedasticity was confirmed by visual inspection of the scatterplot of standardized residuals and standardized predicted values appearing to be randomly scattered. Normal distribution of residuals was assessed by visual inspection of a histogram and a normal probability plot.

Results indicated that higher subjective perceived task demand as measured by the overall NASA TLX significantly predicted pupil dilation, $F(1, 48) = 4.16$, $p = .047$. Specifically, scores on the overall NASA TLX accounted for 8.0% of the variation in pupil dilation, adjusted $R^2 = 0.061$ (see Table 2). Predictions were made to determine pupil diameter for those people who had an overall TLX score of 20, 50, and 80 using the regression equation: pupil dilation = 3.178 + 0.009 x (NASA TLX score; see Figure 12). For a NASA TLX score of 20, pupil dilation was predicted as 3.35 mm, 95% CI [3.16, 3.54]; for a score of 50 it was predicted as 3.61 mm, 95% CI [3.40, 3.82]; and for a score of 80 it was predicted as 3.87 mm, 95% CI [3.44, 4.31].

Table 2

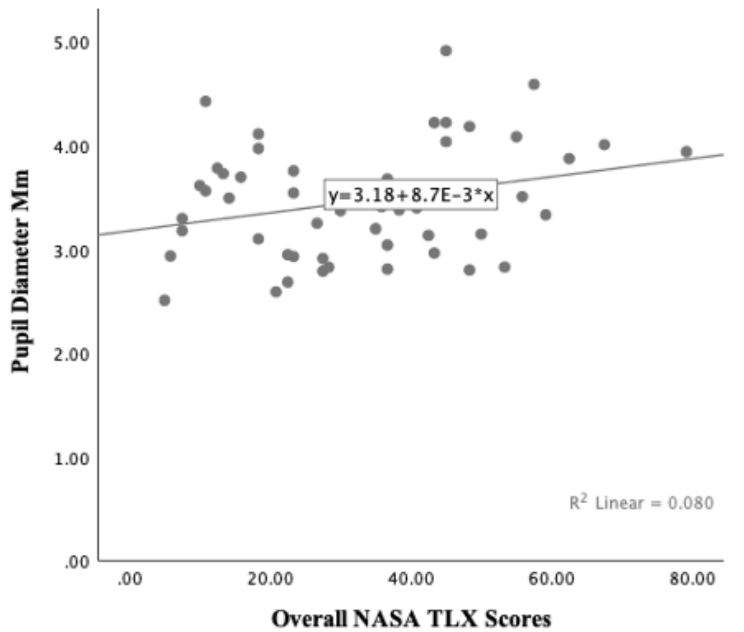
Regression Analysis Summary for NASA TLX predicting Pupil Dilation.

Variable	*B*	95% CI for Mean		*β*	t	*p*
		LL	*UL*			
(Constant)	3.18	2.86	3.96		19.90	.000
Overall TLX	.009	.000	.017	0.47	2.04	.047

Note. R^2 adjusted = 0.061. CI = Confidence interval, LL = lower limit; UL = upper limit.

Figure 14

Relationship between scores on the NASA TLX and pupil diameter in millimeters.



Hypothesis H4. Subjects with larger pupil dilation were expected to have decreased likelihood of IB occurrence compared to those with smaller pupil dilation (Beatty & Lucero-Wagoner, 2000; Dehais et al., 2020). Specifically, subjects with larger pupil dilation were expected to be more likely to detect the critical stimulus than subjects with smaller pupil dilation. A logistic regression was conducted with the pupil diameter and IB.

A previously discussed, there are seven assumptions for logistic regression. The first four are related to study design. The dependent variable is dichotomous (detect or fail to detect), the independent variable is measured on a continuous scale (pupil diameter in mm), independence of observations, and greater than 15 cases in each group. Linearity of the continuous variable with respect to the logit of the dependent variable was confirmed using the Box-Tidwell (1962) procedure, $p > .05$. Outliers were assessed using casewise standardized residuals and none were identified.

A logistic regression analysis was performed to examine the effects of pupil dilation on the likelihood that participants experienced IB (failed to detect the runway incursion). The predictor variable, pupil dilation, in the logistic regression analysis was found to significantly contribute to the model: $B = -1.40$, $SE = .64$, $Wald = 4.80$, $p = .028$ (see Table 3). The model explained 14.5% (Nagelkerke R^2) of the variance. The estimated odds ratio favored an increase of a subject experiencing IB (failing to detect a stimulus) by a factor of 4.03 (1 / .248 = 4.03) for each mm decrease in pupil dilation [*Exp (B)* = .025, 95% CI[0.07, 0.86]. The area under the ROC curve was .694, 95% CI [.542, .847], which is an acceptable level of discrimination.

Table 3

Logistic Regression Predicting Likelihood of Inattentional Blindness by Pupil Dilation.

	β	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Mean	
							LL	*UL*
Pupil Dilation	-1.40	.64	4.80	1	.028	.248	.07	.86
Constant	4.27	2.16	3.90	1	.048	71.36		

Note: Pupil dilation is in mm.

CHAPTER VI

DISCUSSION

Inattentional blindness is the failure of observers to notice the presence of an unexpected but easily viewable event. IB often occurs when cognitive resources are diverted elsewhere. The present research examined how subjective level of task load, the degree of automation (fully automated, partial automation, and manual), and pupil dilation were associated with IB in a simulated flight landing task.

Hypothesis 1 examined whether subjective task demand scores on the NASA-Task Load Index (TLX) would vary significantly as function of automation condition. (Driskell, Driskell, & Salas, 2013; Hancock & Warm, 1989; Hart & Staveland, 1988; Recarte et al., 2008). Based on previous research (Driskell, Driskell, & Salas, 2013; Hancock & Warm, 1989; Hart & Staveland, 1988; Recarte et al., 2008), it was expected that subjects in the full automation condition would report the lowest overall task demand, subjects in the partially automated condition would report higher task demand than the automated condition but lower task demand than the manual condition, and subjects in the manual condition would report the highest overall task demand.

Partial support was found for Hypothesis 1. That is, overall NASA-TLX scores were significantly lower for the full automation condition as compared to the partial automation condition. Although in the expected direction, scores for the partial automation condition did not significantly differ from the manual condition. This finding is in line with a visual search impairment study conducted by Recarte et al. (2008) that used scores on the NASA-TLX to compare subjective perception of task load during performance of combinations of tasks that were visually demanding and/or cognitively demanding resulting in a $\eta^2 = .152$ for the combined visual demand and cognitive task.

In particular, findings from the present study reflect both visual demand and cognitive task, both of which vary by automation demands. The full automation task included monitoring the aircraft automation during an automated landing. The subject monitored that the aircraft maintained targeted speeds, reduced speed at a target altitude, and landed on the runway. Successful completion of this task required very little cognitive load which was confirmed by the data. The manual automation condition required physical control of both speed and heading in light winds through the same speed change at target altitude and to landing on the correct runway. The successful completion of the manual task necessitated higher workload to complete the task which was shown in the data. The partial automation condition required physical control of the heading but the speed was automated. This task necessitated moderate workload which was confirmed by the data. This finding also serves as a manipulation check that the automation conditions were subjectively experienced as low, moderate, and high task load.

Results of Hypothesis 2 confirmed that the failure to detect the critical stimulus was significantly related to the automation condition. As predicted, the partial automation condition had the lowest likelihood of IB despite the full automation condition exhibiting the lowest task load. The increased likelihood of IB in full automation (low task load) within a similar range of manual operation (high task load) indicates that high task load should not be the only hazardous state of awareness considered in task planning. Ultimately, 12 (60%) subjects in the full automation condition failed to detect the critical stimulus compared to 4 (20%) subjects in the partial automation condition and 8 (40%) subjects in the manual condition.

These results have practical implications. High task loading has a predictable detriment to task performance. In addition, low task loading also presents a concern. To aim for a task load that maintains a balance between too much and not enough, a method for assessing human operator state during task operation is needed to both optimize tasking and to determine if the human has entered into a hazardous state of awareness. Stephens et al. (2018) provided a review on the use of biocybernetics loops for use in adaptive automation implementation. Physiologically adaptive automation involves detection of a transient cognitive state that induces a modification of a functional aspect of an external system such as triggering a warning or advanced controls system of an aircraft that then interacts or modifies with the human cognitive state (Stephens et al., 2018).

It was hypothesized that subjects who reported higher task demand (i.e., reported higher scores on the overall NASA TLX) were expected to have increased pupil dilation as compared to subjects who reported low task demand (Recarte et al., 2008). Results indicated that subjects who indicated a higher task demand on the NASA TLX exhibited significantly increased pupil dilation as compared to subjects who indicated a lower task demand; however, the effect size was quite low, accounting for 8.0% of the variation in pupil dilation. This relatively low level of variance accounted for may be because the study design permitted subjects to move their head freely to increase realism of flight task performance, which may have increased the amount of noise in the eye tracking data. In addition, the author approached the pupil dilation data cleaning and removal of participant data conservatively to maintain sufficient sample size. Despite the low effect size, this finding is significant and demonstrates that the subjective workload documented on the TLX was in agreement with the pupil dilation in the expected direction. Specifically, participants who experience the task as high demand also exhibited increased pupil dilation. An important advantage of pupillary response over retrospective subjective report, i.e., the NASA TLX, as a predictor of IB occurrence is that pupil dilation can be measured passively and in real-time.

Hypothesis 4 explored if subjects with larger pupil dilation would exhibit a decreased likelihood of IB occurrence compared to those with smaller pupil dilation (Beatty & Lucero-Wagoner, 2000; Dehais et al., 2020). The current study found support for this hypothesis and success using pupil dilation as a method to distinguish between individuals who exhibited IB (failed to detect the runway incursion) and those who did not. That is, subjects with larger pupil dilation had a decreased likelihood of IB occurrence compared to those with smaller pupil dilation. Specifically, those subjects with larger pupil dilation were more likely to detect the critical stimulus than those subjects with smaller pupil dilation. The relationship between the pupil dilation and IB was significant and explained 14.5% of the variance in IB.

The established methodology for assessing the occurrence of IB is retrospective subjective reports (Mack & Rock, 1998; Simons & Levine, 1999). As discussed in the examples of IB occurring in daily life, collecting the statements after an accident can explain what happened but cannot intercede in real time. In contrast, pupil dilation is a physiological response. The size of the pupil has been shown to indicate cognitive state, with larger dilation indicating increased workload. As expected, those participants with larger pupil dilation were more likely to detect the runway incursion. These findings are promising as they suggest that pupil dilation may be another tool to assist in understanding IB. This study supported the potential for utilizing a non-invasive, passive observation system in the form of an eye-tracker to gain real-time information related to pilot state via pupil dilation to identify periods of increased potential for an IB occurrence.

Limitations and Future Research

This study utilized non-pilots completing a complex task which was a version of a simulated aviation task designed for non-pilots. A future study should engage pilots in a realistic flight simulation to examine pupil dilation and IB to runway incursions. Attention should be paid to the design of the simulation to include a pilot subject matter expert to ensure the realism and plausibility of both the increased cognitive workload and the runway incursion. Elements of nominal task performance (e.g., complex approach plates, emergencies, unexpected equipment failures, and radio communications) can be combined to produce a realistic and believable scenario with increased complexity sufficient to increase the likelihood of an IB event. Related to the sample limitations, race representativeness should be considered in future studies.

The difficulties working with pupil dilation data are not inconsequential. Unlike the majority of pupil dilation studies, the subjects in this study were permitted to perform the flight task with free movement of their head and eyes in an attempt to maintain relevance to the applied task. This free movement increased the amount of noise in the data, despite capitalizing on the state eye tracking at the time. The author chose a conservative approach to removing pupil dilation data. Future studies with free motion designs should increase the sample size to allow for more liberal removal of subject data with non-optimal pupil dilation data. The restriction of head movement is one option but if the exploration of IB occurrences during an operational task is the key interest, the movement restrictions may reduce the generalizability of the IB findings.

In addition, eye tracking technology has increased exponentially even in the short time since data for this study were collection. Aided by improvements in miniaturization and battery technology, issues such as head angle, tracking model, exiting the tracking box, and an obscured view are now longer a concern for the smaller, faster, head-worn eye trackers. Cumbersome external computers and time-intensive calibration (as detailed in Appendix J) are also no longer a problem with these next generation eye tracking systems now available. For example, the eye tracker used in this system contained four cameras and two infrared flashers mounted in fixed locations in front of the subject. This version of the eye tracker required detection of the eye in two of the four cameras. With this setup, subject data were unusable if the subject tilted their head too much or moved out of the head box which could happen by slouching over time or leaning very far forward. Current head-worn systems have high benefit against studies permitting free range of movement. Several data quality issues experienced in this study would be greatly reduced if the study were replicated using a head-worn glasses-style eye tracker. Future research should be conducted with updated technology and an increased sample size to permit continued freedom of head/eye movement and enable liberal data cleaning criteria. Nevertheless, despite these limitations, the results from this study suggest pupil dilation may be a method for successfully predicting periods apt for increased IB occurrences.

CHAPTER VII

CONCLUSIONS

Deadly consequences can occur when an operator fails to detect to an object or event relevant to task performance, but little capability exists regarding the real-time detection of IB prior to an incident. Understanding the conditions that increase the likelihood of a person to have an IB occurrence can facilitate the development of real-time mitigation strategies. Scores on the NASA-TLX varied significantly by automation condition with the full automation having the lowest subjective task load followed by partial automation and then manual with the highest. IB detection varied significantly across automation condition; however, the moderate workload condition, in which there was partial automation, exhibited the lowest likelihood of IB occurrence. The low workload full automation condition had a similar likelihood of IB as the manual condition.

Eye tracking is one psychophysiological assessment technique that is unobtrusive and can operate passively. Eye tracking can be deployed to monitor an operator *in situ* to identify periods of increased likelihood for attention lapses. Based on the neurophysiological conceptualization of attention and IB by Dehais et al. (2020), the current study examined pupil dilation as a method to differentiate between those who detected a runway incursion and those who did not during a simulated flight landing task across three levels of automation. Subjects who reported higher task demand had increased pupil dilation and subjects with larger pupil dilation were more likely to detect the runway incursion. Support was found that pupil dilation was able to significantly discriminate the detectors from those who failed to detect. Specifically, those who detected the critical stimulus had increased pupil diameter as compared to those who failed to detect. This final result suggests eye tracking may provide a real-time IB mitigation opportunity to identify when unexpected visual stimulus detection is reduced.

REFERENCES

Ahlström, C., Kircher, K., & Kircher, A. (2009). Considerations when calculating percent road centre from eye movement data in driver distraction monitoring. *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 132-139. https://doi.org/fd2q

Australian Transport Safety Bureau (ATSB) (2012). *Aircraft proximity event – two Cessna 172S, VH-EWE and VH-EOP, Moorabbin Airport, Victoria.* Aviation Occurrence Investigation [AO-2012-099]. https://www.atsb.gov.au/media/4027309/AO-2012-099%20Final.pdf

Bailey, N. R. & Scerbo, M. W. (2007) Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science, 8*(4), 321-348. https://doi.org/d4r96d

Beanland, V. & Pammer, K. (2011). Failures of visual awareness: Inattentional blindness and attentional blink. *i-Perception, 2*(4), 227-227. https://doi.org/10.1068/ic227

Beanland, V., & Chan, E. H. C. (2016). The relationship between sustained inattentional blindness and working memory capacity. *Attention, Perception, & Psychophysics, 78*(3), 808-817. https://doi.org/10.3758/s13414-015-1027-x

Becklen, R., & Cervone D. (1983). Selective looking and the noticing of unexpected events. *Memory & Cognition, 11*(6), 601-608. https://doi.org/10.3758/BF03198284

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

Bressan, P., & Pizzighello, S. (2008). The attentional cost of inattentional blindness. *Cognition, 106*(1), 370-383. https://doi.org/b8rsx2

Broadbent, D. (1958). *Perception and Communication.* Pergamon Press.

Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control, *Quarterly Journal of Experimental Psychology*, *29*(2), 181-201. https://doi.org/fmvxg6

Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition, 102*(3), 321-340. https://doi.org/d7xft4

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*, 975-979. https://doi.org/d4m87v

Court of Inquiry (COI) 2001. Daily Transcripts. US Navy record of proceedings of a Court of Inquiry convened at Trial Service Office Pacific by order of Commander In Chief United States Pacific Fleet to inquire into a collision between USS Greeneville (SSN 772) and Japanese M/V Ehime Maru that occurred off the coast of Oahu, Hawaii on 9 February 2001. Naval Station Pearl Harbor, Hawaii. https://www.jag.navy.mil/library/investigations/GREENEVILLE_Combined_COI_Rpt.pdf

De Bittencourt, P. R. M. (2004). Photosensitivity: The magnitude of the problem. *Epilepsia, 45*(1), 30-34. https://doi.org/b6nk4r

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: Evidence for inattentional deafness. *Human Factors*, *56*(4), 631-644. https://doi.org/fd26

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in Neuroscience*, *14*, 268. https://doi.org/gg955z

Driskell, T., Driskell, J. E., & Salas, E. (2013). Stress, performance, and decision making in organizations. In S. Highhouse, R.S. Dalal, & E. Salas (Eds.), *Judgment and Decision Making at Work* (pp. 251-276). Routledge.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 271-285. https://doi.org/fpf9qx

Ellis, K. K. E. (2009) *Eye tracking metrics for workload estimation in flight deck operations* [Unpublished master's thesis]. University of Iowa. https://doi.org/10.17077/etd.a773626l

Fairclough, S. H., Gilleade, K., Ewing, K. C. & Roberts, J. (2013). Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems, 6*(1), 63–79. https://doi.org/fd3c

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/bxjdcg

Federal Aviation Administration. (2012). *Runway Safety Report 2011-2012* (FAA Publication No. 2012-AJS-475). http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/safety/media/2012-AJS-475-FY2011-Runway-Safety-Annual-Report.pdf

Federal Aviation Administration. (2020). *Runway Incursion Totals by quarter FY2019 vs. FY2018*. https://www.faa.gov/airports/runway_safety/statistics/year/?fy1=2019&fy2=2018

Felsten, G., & Wasserman, G. S. (1980). Visual masking: Mechanisms and theories. *Psychological Bulletin, 88*(2), 329-353.

Finomore Jr, V. S., Shaw, T. H., Warm, J. S., Matthews, G., & Boles, D. B. (2013). Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Human Factors*, *55*(6), 1044-1063. https://doi.org/10.1177/0018720813484498

Fischer, E., Haines, R. F., & Price, T. A. (1980). *Cognitive issues in head-up displays (*NASA Technical Report No. 1711). NASA Ames Research Center.

Flägel, K., Galler, B., Steinhäuser, J., & Götz, K. (2019). The NASA-TLX - an instrument for measuring consultation workload within general practice: Evaluation of psychometric properties. *The Journal of Evidence and Quality in Health Care, 147-148,* 90–96. https://doi.org/10.1016/j.zefq.2019.10.003

Fraher, A.L. (2010). The USS Greeneville submarine collision: A breakdown in team resource management (TRM). In G. Graen (Ed.), *LMX leadership series: Leadership in harm's way* (pp. 1-36). Information Age Publishing. http://www.amyfraher.com/uploads/USS_Greeneville-TRM_Breakdown-Fraher-2010.pdf

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. *Quarterly Journal of Experimental Psychology, 66*(12), 289-2294. https://doi.org/fd7g

Hancock, P. A. & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519-537. https://doi.org/10.1177/001872088903100503

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. https://doi.org/fzvtd4

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati, *Human mental workload* (pp. 139-183). North-Holland.

Heffernan, M. (2011). *Wilful [sic] Blindness: Why we ignore the obvious at our peril.* Bloomsbury USA.

Hutchinson, B. (2019). Toward a theory of consciousness: A review of the neural correlates of inattentional blindness. *Neuroscience and Biobehavioral Reviews, 104*, 87-99. https://doi.org/fd7p

Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2010). Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology, 24*(5), 597-607. https://doi.org/bfzk76

Jensen, M. S., Yao, R., Street, W. N., & Simons, D. J. (2011). Change blindness and inattentional blindness. *Cognitive Science, 2*(5), 529-546. https://doi.org/10.1002/wcs.130.

Kahneman, D. (1973). *Attention and effort*. Prentice Hall.

Kahneman, D. & Peavler, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology, 79*(2), 312-318. https://doi.org/d3h76c

Kanwisher, N. G., & Potter, M. C. (1990). Repetition blindness: Levels of processing. *Journal of Experimental Psychology Human Perception and Performance, 16*(1), 30-47.

Kennedy, K. D. & Bliss, J. P. (2013). Inattentional blindness in a simulated driving task. *Proceedings of the 57th Human Factors and Ergonomics Society Annual Meeting*, *(57)*1, 1899-1903. https://doi.org/fd7r

Kennedy, K. D., Stephens, C. L., Williams, R. A., & Schutte, P. (2014). Automation and inattentional blindness in a simulated flight task. *Proceedings of the 58th Annual Meeting of the Human Factors and Ergonomics Society*, *58*(1), 2058-2062. https://doi.org/fd7s

Kennedy, K. D., Stephens, C. L., Williams, R. A. & Schutte, P. (2017). Repeated induction of inattentional blindness in a simulated aviation context. *Proceedings of the 61st Annual Meeting of the Human Factors and Ergonomics Society, 61*(1), 1959-1963. https://doi.org/fd7t

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6*(3), 120-125. https://doi.org/b7kbhn

Koivisto, M., Hyönä, J., & Revonsuo, A. (2004). The effects of eye movements, spatial attention, and stimulus features on inattentional blindness. *Vision Research*, *44*(27), 3211-3221. https://doi.org/c42r3r

Kuhn, G., & Findlay, J. M. (2010). Misdirection, attention and awareness: Inattentional blindness reveals temporal relationship between eye movements and visual awareness. *Quarterly Journal of Experimental Psychology, 63*(1), 136-146. https://doi.org/bc2p48

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effects of speech-based e-mail on drivers' attention to the roadway. *Human Factors*, *43*(4), 631-640. https://doi.org/ckhhx4

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. https://doi.org/dr6jf9

Lum, T. E., Fairbanks, R. J., Pennington, E. C., & Zwemer, F. L. (2005). Profiles in patient safety: Misplaced femoral line guidewire and multiple failures to detect the foreign body on chest radiography. *Academic Emergency Medicine, 12*(7), 658-662. https://doi.org/cnzsm3

Mack, A., & Rock, I. (1998). *Inattentional blindness*. The MIT Press.

Mack, A., & Rock, I. (1999). Inattentional blindness an overview by Arien Mack and Irvin Rock. *Psyche,* 28, 1059 –1074. http://journalpsyche.org/files/0xaa59.pdf

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1). http://doi.org/10.5334/joc.18

Memmert, D. (2006). The effects of eye movements, age, and expertise on inattentional blindness. *Consciousness and Cognition, 15*(3), 620–627. https://doi.org/dt9323

Mohan, D. B., Shree, J., Prabhakar, G., Saluja, K. S., Pashilkar, A., & Biswas, P. (2019). Estimating pilots' cognitive load from ocular parameters through simulation and in-flight studies. *Journal of Eye Movement Research, 12*(3). 1-16, https://doi.org/fd7x

Most, S. B., & Astur, R. S. (2007). Feature-based attentional set as a cause of traffic accidents. *Visual Cognition, 15*(2), 125-132. https://doi.org/10.1080/13506280600959316

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review, 112*(1), 217-242. https://doi.org/d572jd

Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological Science, 12*(1), 9–17. https://doi.org/fmbz72

National Transportation Safety Board (NTSB). (2001). Accident Report No. (DCA-01-MM-022). https://ntsb.gov/investigations/AccidentReports/Reports/MAB0501.pdf

National Transportation Safety Board (NTSB). (2007). Airport runway accidents, serious incidents, recommendations, and statistics. http://www3.alpa.org/portals/alpa/runwaysafety/NTSBRunwaySafetyfactsheet.pdf

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of Human Factors and Ergonomics, 7*(1), 58-99. https://doi.org/drztnr

Neigel, A. R., Claypoole, V. L., Smith, S. L., Waldfogle, G. E., Fraulini, N. W., Hancock, G. M., Helton, W. S., & Szalma, J. L. (2020). Engaging the human operator: A review of the theoretical support for the vigilance decrement and a discussion of practical applications. *Theoretical Issues in Ergonomics Science*, *21*(2), 239-258. https://doi.org/fd7z

Neisser, U. (1979). The control of information pickup in selective looking. In A.D. Pick (Ed.). Perception and its development: A tribute to Eleanor J. Gibson (pp 201-219). Erlbaum.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*(4), 480–494. https://doi.org/frh4zc

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522-536. https://doi.org/10.1037/h0026699

Oktay, B., & Cangöz, B. (2018). I thought I saw "Zorro": An inattentional blindness study. *Archives of Neuropsychiatry*, *55*(1), 59. https://doi.org/fd72

Parasuraman, R., Cosenzo, K., & de Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21,* 270–297. https://doi: 10.1080/08995600902768800

Parasuraman, R. & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381-410. https://doi.org/fcmx4h

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253. https://doi.org/bbntx8

Parasuraman, R., & Rizzo, M. (2008). *Neuroergonomics: The brain at work* [1st Ed]. Oxford University Press.

Peißl, S., Wickens, C.D., & Baruah, R. (2018). Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology, 28*(3-4), 98-112, https://doi.org/10.1080/24721840.2018.1514978

Pizzighello, S., & Bressan, P. (2008). Auditory attention causes visual inattentional blindness. *Perception, 37*(6), 859-866. https://doi.org/10.1068/p5723.

Pope, A. T., Stephens, C. L., & Gilleade, K. (2014). Biocybernetic adaptation as biofeedback training method. In S. H. Fairclough & K. Gilleade (Eds.), *Advances in Physiological Computing, Human–Computer Interaction Series* (pp. 91–115). Springer. https://doi.org/10.1007/978-1-4471-6392-3_5

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25. https://doi.org/10.1080/00335558008248231

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology, 109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*(1), 31-43. https://doi.org/c6kd37

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied, 9*(2), 119–137. https://doi.org/dbfp2k

Recarte, M. A., Perez, E., Conchillo, A., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink and subjective rating. *Spanish Journal of Psychology*, *11*(2), 374-385. https://doi.org/10.1017/S1138741600004406

Reyes, M. L., & Lee, J. D. (2008). Effects of cognitive load presence and duration on driver eye movements and event detection performance. *Transportation Research Part F: Traffic Psychology and Behaviour, 11*(6), 391-402. https://doi.org/10.1016/j.trf.2008.03.004

Richards, A., Hannon, E. M., & Vitkovitch, M. (2012). Distracted by distractors: Eye movements in a dynamic inattentional blindness task. *Consciousness and Cognition, 21*(1), 170-176. https://doi.org/10.1016/j.concog.2011.09.013

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology, 53*(1), 61-86. https://doi.org/bfx2qt

Scerbo, M. W., Freeman, F. G., Mikulka, P. J., Parasuraman, R., Di Nocera, F., & Prinzel, L. J. (2001). *The efficacy of psychophysiological measures for implementing adaptive technology* (Technical Paper No. 211018). NASA Langley Research Center

Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2003). A brain-based system for adaptive automation. *Theoretical Issues in Ergonomic Science, 4*(1-1), 200–219. https://doi.org/ckdfxp

Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences, 1*(8), 291-296. https://doi.org/10.1016/S1364-6613(97)01094-2

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences, 4*(4), 147-155. https://doi.org/10.1016/S1364-6613(00)01455-8

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059–1074. https://doi.org/10.1068/p281059

Simons, D. J., & Jensen, M. S. (2009). The effects of individual differences and task difficulty on inattentional blindness. *Psychonomic Bulletin & Review, 16*(2), 398-403. https://doi.org/cfm49q

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267. https://doi.org/bq9qc7

Singh, N. (2009). On a wing and a prayer: surgeons learning from the aviation industry. *Journal of the Royal Society of Medicine*, *102*(9), 360-364. https://doi.org/10.1258/jrsm.2009.090139

Smart Eye. (2008). *Smart Eye Pro 5.0 - User Manual* (Rev. 249). Gothenburg, Sweden.

Stephens, C., Dehais, F., Roy, R. N., Harrivel, A., Last, M. C., Kennedy, K., & Pope, A. (2018). Biocybernetic adaptation strategies: machine awareness of human engagement for improved operational performance. In *International Conference on Augmented Cognition*, (pp. 89-98). https://doi.org/10.1007/978-3-319-91470-1_9

Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science, 16*(3), 128-131. https://doi.org/10.1111/j.1467-8721.2007.00489.x

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology Applied, 9*(1), 23-32. https://doi.org/10.1037/1076-898X.9.1.23

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th Edition). Pearson Education.

Transport Accident Investigation Commission (TAIC). (2010). *Final Report: Runway Incursion, Dunedin International Airport* (TAIC Aviation Inquiry 10-006). http://www.taic.org.nz/ReportsandSafetyRecs/AviationReports/tabid/78/ctl/Detail/mid/482/InvNumber/2010-006/language/en-US/Default.aspx

Treisman, A. M., 1964. Selective attention in man. *British Medical Bulletin, 20*(1)*,* 12-16. https://doi.org/10.1093/oxfordjournals.bmb.a070274

Treisman A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/gwh

Varakin, D. A., Levin, D. T., & Fidler, R. (2004). Unseen and unaware: Implications of recent research on failures of visual awareness for human-computer interface design. *Human-Computer Interaction, 19*(4), 389-422. https://doi.org/10.1207/s15327051hci1904_9

Wayand, J., & Levin, D. T. (2001). Ignoring a merciless act [Abstract]. *Journal of Vision, 1*(3) http://journalofvision.org/1/3/12/, doi:10.1167/1.3.12

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D., & Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *The International Journal of Aviation Psychology*, *19*(2), 182-199. https://doi.org/10.1080/10508410902766549

Wickens, C. D., & Carswell, C. M. (2012). Information Processing. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics.* https://doi.org/10.1002/9781118131350.ch5

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18. https://doi.org/10.1207/S15327108IJAP1201_2

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing - Pupillometry in Hearing Science, 22*. https://doi.org/10.1177/2331216518800869

Wright, T. J., Boot, W. R., & Morgan, C. S. (2013). Pupillary response predicts multiple object tracking load, error rate, and conscientiousness, but not inattentional blindness. *Acta Psychologica, 144*(1), 6-11. https://doi.org/10.1016/j.actpsy.2013.04.018

Wood, K., & Simons, D. J. (2019). Now or never: Noticing occurs early in sustained inattentional blindness. *Royal Society Open Science*, *6*(11). https://doi.org/10.1098/rsos.191333

Young, M. S. & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors, 44*(3), 365-375. https://doi.org/10.1518/0018720024497709

APPENDIX A

HUMAN SUBJECTS INFORMED CONSENT STATEMENT

Subject:	Condition: M P A	Date:	Start Time:	End Time:

Title of Research: "A study of inattentional blindness and steady state visually evoked potentials using cortical/physiological and self-report measures."

Principal Investigators:

Kellie Kennedy	NASA LaRC	757-XXX-XXXX
Ralph Williams	NASA LaRC	757-XXX-XXXX
Chad Stephens	NASA LaRC	757-XXX-XXXX
Alan Pope	NASA LaRC	757-XXX-XXXX

Federal regulations require researchers to obtain signed consent for participation in research involving human subjects. After reading the information and the Statement of Consent below, if you wish to consent, please indicate so by signing this form.

I. Statement of Procedure:

Thank you for your interest in this research. Experimental rationale and procedures have been discussed with you in detail. You will find a summary of the major aspects of the test and associated research, including the risks and benefits of participating in the following sections.

Please read the following information carefully. If you wish to participate in this study, sign your name and date the form in the space provided. Any information you provide will be kept in strict confidence to protect your privacy

II. I understand that:

- This is a research experiment and I will be one of approximately 20 subjects.
- I understand that screening data (including self-reported health status) will be collected and stored confidentially, scored, and used by the Principal Investigators to determine whether I am selected to participate in the experiment.
- I may voluntarily discontinue or be asked to discontinue participation in this study at any time without penalty or loss of benefits to which I am otherwise entitled.
- This experiment will be performed in the Human and Autonomous Vehicle Systems (HAVS) laboratory facility (in building XXXX) at NASA Langley Research Center.
- The duration of my participation will include one session lasting approximately 1 hour. I may take a break at any time, though I am encouraged to complete each scenario before taking a break.
- I will be participating in an experiment designed to identify the parameters of hazardous states of awareness and steady state visually evoked potential.
- I will receive a briefing of and training for the operation of the equipment. I will be allowed time to familiarize myself with the software/equipment prior to starting the experiment. I will also participate in a debriefing at the end of the experimental session.
- I will be asked to perform a variety of simulated instrument procedures in a part-task, fixed-base simulator. These procedures will reflect current and potential future operations.
- During the course of the experiment, I will provide my impressions and assessments by providing verbal inputs, as well as completing written/computerized questionnaires and surveys.

- The evaluation scenarios may include non-normal or failure conditions in addition to normal operation. These non-normal situations may not be announced prior to the evaluation to avoid prejudice or expectation on my part.
- Records of my participation will be kept confidential by encoding them with a subject identification number.
- Prior to the start of the experimental task I will be connected and calibrated to physiological recording equipment: electroencephalogram (EEG), electrocardiogram (ECG), and respiration (RESP).
- My physiological responses to the task as measured by EEG, ECG, and RESP will be monitored and recorded.
- I understand that the data files recorded during my participation in this experiment will be shared with other researchers and that these files will be identified only by the subject number assigned by the experimenter. I do voluntarily consent to sharing the data files recorded during my data collection session, as long as my identity is not disclosed. Furthermore, the results identified by subject number may be published in the form of conference papers, journal articles, and formal NASA reports.
- A video and audio recording of my person during the session will be made with a closed-circuit video camera for post hoc behavioral analysis. The video recording is intended to provide a visual record of my interaction with the automation interfaces. The video will not be released or shown to anyone other than the Principal Investigators identified above except it will be shared as outlined below, but only if I grant my explicit permission by signing “Approve”, as set forth below. This permission is in addition to my consent to participate in this research that I will grant by signing on this form. The audio record may be distributed to other researchers in addition to the Principal Investigators or published in reports and shared with others outside NASA after it has been transcribed to text and any personally identifying remarks or information that may be associated with me will be removed.
- Eye tracking/pupillometry (ET) data will be collected during this experiment. The SmartEye Eye Tracking system is a non-evasive infrared (IR) camera-based eye tracking unit. This system will record both still and video images of my person while participating in this study.
- I consent to allow still and/or moving images of my person captured with a closed-circuit video camera and the SmartEye Eye Tracking system and audio recordings of my voice to be shared with other researchers within in outside NASA during analysis and disseminated in reports of this work. These other researchers are in addition to the Principal Investigators identified above. I understand that if I approve sharing still and/or moving images of my person with other researchers during analysis and dissemination of still and moving images of my person and audio recordings of my voice in published reports of this work, my participation in this research will no longer be anonymous and someone with whom the still and or moving images are shared may recognize me and, whether or not I am recognized, my participation in this research will be known to others in addition to the Principal Investigators identified above.

(Sign One ONLY):

APPROVE ______________________________	**DISAPPROVE** ______________________________
Volunteer Subject Signature	Volunteer Subject Signature
Signing “Approve” indicates that **I consent** to have still or video images of my person and audio of my voice recorded and shared with other researchers outside and within NASA during analysis and disseminated in reports of this work. I understand that the other researchers are in addition to the Principal Investigators identified above.	Signing “Disapprove” indicates that **I do not consent** to sharing still or moving video images of my person or audio recordings of my voice or any other potentially personally identifying information with anyone other than the Principal Investigators identified above. I do not consent to still or moving video images of my person or audio recordings of my voice in text being disseminated in reports of this work.

- All data recorded during this experiment will be stored under lock and key in a different filing cabinet from informed consent forms in the Building XXXX, Room XXXX. All electronic data and still and moving video images and audio recordings of my voice recorded during this experiment will be saved in a manner that does not associate me with the data, still images, or video images, unless I approve of release of such data, as set forth above. It will then be stored on a password protected computer in Building XXXX, Room XXX. The password will only be known by the Principal Investigators.
- I may contact the investigators listed above if I have any questions regarding this experiment before, during, or after my participation.

III. Confidentiality:

Records of my participation will be kept confidential by encoding them with subject identification numbers. Any data published as a result of this research will not include any personally identifiable information that could be linked to or associated with me; however, if I have approved of release of still and or moving video images of my person and audio recordings of my voice, as described above, it is possible someone will recognize my person, and, whether or not someone recognizes me, such images and recordings will associate my likeness with having participated in this research.

IV. Compensation

Civil servant volunteers who participate in the research do so in their official capacity. A civil servant injured during the course of this research may file for compensation through the Federal Workers Compensation System. For additional information, participants may contact the LaRC Office of Human Resources at 757-XXX-XXXX.

As a non-civil servant, I will receive no form of compensation.

For non-civil servants, insurance coverage is provided to each research subject volunteer under the NASA Langley Teams II contract. For additional information, I may contact X at 757-XXX-XXXX. Non-civil servant volunteers injured as a result of participating in the research may also file a claim under the Federal Tort Claims Act by filing Standard Form 95. For additional information, participants may contact the LaRC Office of Chief Counsel at 757-XXX-XXXX.

V. Potential Risks

- Participating in this research will not create any foreseeable risks to my health. No physical discomforts are expected in this test other than those normally associated with operating a fixed-based simulator, such as fatigue or eye strain.
- An aspect of this research employs flashing lights in the form of strobe lights. Some individuals with photosensitive epilepsy are known to suffer seizures when exposed to certain flashing lights; however, photosensitive epilepsy is most common in children and adolescents under age 20. Nevertheless, in very rare instances, individuals with no history of epilepsy have suffered seizures when exposed to flashing lights like strobe lights.
- In the unlikely event that you are injured or otherwise experience discomfort while at NASA Langley, you may visit the on-site Occupational Health Clinic. The Clinic has hours of operation from 7:00 a.m. to 3:00 p.m. The clinic number is 757-XXX-XXXX. Emergency medical personnel and ambulance service is also available to transport you to nearby health care providers.
- If you have questions about the research and your rights should you experience any injury, you may contact the principal investigators listed at the beginning of this document.

VI. Potential Benefits

- You will derive no direct benefit from your participation in this study.
- The results of your participation may improve the safety of commercial air travel for a broad class of aircraft.

VII. Voluntary Participation

Taking part in this study is voluntary. You may withdraw from participating or be asked to withdraw from participating at any time. Such a decision that will not result in any penalty or loss of benefits to which you may otherwise be entitled.

VIII. Safety

As a voluntary test subject participating in this research, I understand that:

- NASA is committed to ensuring my safety, health, and welfare plus the safety and health of all others involved with this research.
- I should report any accident, injury, illness, and changes in my health condition, hazards, safety concerns, or health concerns to the above listed investigators. If I am unable to reach the above named individuals or am not satisfied with the response I receive, I should contact the LaRC Safety Office at 757-XXX-XXXX or the Chairperson of the LaRC Institutional Review Board, Mr. XXX XXXXXXXXX, at 757-XXX-XXXX.
- If I detect any unsafe condition that presents an imminent danger to me, or others, I have the right and authority to stop the activity or test. In such cases the Principal Investigator and associated research personnel will comply with my direction, stop the activity, and take action to address the imminent danger.

IX. Statement of Consent:

- I certify that I have read and fully understand the explanation of procedures, benefits, and risks associated with the research herein, and I agree to participate in the research described herein. My participation is given voluntarily and without coercion or undue influence, and I also voluntarily consent to sharing the data files recorded during my data collection session, as long as my identity is not disclosed. I understand that I may discontinue participation at any time. I have been provided a copy of this consent statement. If I have any questions or modifications to this consent statement, they are written below.

____
Participant Printed Name

____
Participant Phone Number

____
Participant Street Address

____
Participant City, State, & ZIP

____
Participant Signature   Date

____
Witness Signature   Date

__ Participant has been provided with a Privacy Act Statement meeting the requirements outlined in 14 CFR 1212.602

APPENDIX B

DEMOGRAPHIC FORM

Background Questionnaire

INSTRUCTIONS:
It is necessary for us to obtain a very brief background health history in order to determine if you are eligible for participation in the second part of the study. It is very important that you be completely honest. This screening information will be stored and maintained confidentially in the Human Subject Research File associated with this research and not shared with anyone other than the Principal Investigators named on the research and Dr. Conage or clinic or medical staff, as necessary.

1. What is your age, sex, height, and weight?
    a. Age: _____ years
    b. Sex: □ Male □ Female
    c. Height: _____ feet _____ inches
    d. Weight: _____ pounds

2. Are you a pilot? □ No □ Yes
3. If Yes:
    a. How many flight hours do you have total? __________ (Approximately)
    b. How many flight hours in the last 6 months? __________ (Approximately)
    c. How long have you been flying? _____ Years
    d. Please list the aircraft rating(s) you have?

---

4. In the past year, on average, how often do you use a computer? ______ hours per day

5. In the past year, on average, how often do you play computer games on a PC?
    a. ______ days per month
    b. Typically ______ hours at a time.

6. What percent of that time you play video games is a game that involves simulated flight?
    a. ______% of the time.

7. Have you ever experienced:
    a. A concussion? □ N □ Y
    b. Lost consciousness due to a blow to the head? □ N □ Y
    c. Suffered a traumatic brain injury? □ N □ Y
    d. If YES to any of the above, please explain:

---

8. Do you currently have or have you ever had a hearing problem? □ N □ Y
   If Yes, please explain:

---

9. Do you use vision correction? □ N □ Y
10. If yes:
    a. Do you use:
        □ eyeglasses □ contact lenses □ both
    b. Will you use vision correction during the task?
        i. □ eyeglasses □ contact lenses □ neither
    c. With vision correction, do you have any visual acuity problems? □ N □ Y
        i. If Yes, please explain:

____

____

11. Do you have any known color vision problems? □ N □ Y
    a. If Yes, please explain:

____

____

12. Do you currently have or have you ever had any of the following?
    Check Yes or No:
    a. Epilepsy or seizures □ N □ Y
    b. Problems with flashing lights □ N □ Y
    c. Strong reaction to cold weather □ N □ Y
    d. Circulatory problems □ N □ Y
    e. Low blood pressure (Hypotension) □ N □ Y
    f. High blood pressure (Hypertension) □ N □ Y
    g. Cardiovascular disorder/disease □ N □ Y
    h. Stroke □ N □ Y
    i. Asthma □ N □ Y
    j. Lung problems □ N □ Y
    k. Diabetes □ N □ Y
    l. Neurological problems □ N □ Y

    m. If you have checked YES for any of the above conditions, please explain:

____

____

____

____

13. Have you ever been diagnosed as having:
    a. Learning deficiency or disorder □ N □ Y
    b. Reading deficiency or disorder □ N □ Y
    c. Attention deficit disorder □ N □ Y
    d. Attention deficit hyperactivity disorder □ N □ Y

14. List any over-the-counter or prescription medications you are currently taking:

____

____

15. Do you have or have you ever had any other medical conditions that weren't listed?

□ N □ Y

a. If Yes, please explain:

____

____

16. Do you use tobacco products of any kind? □ N □ Y

17. If Yes:

a. What kind? ___________________________

b. How often do you use this product? ____________

c. When was the last time you used this product? _____________

18. Do you consume caffeine? □ N □ Y

19. If yes, what is your average daily caffeine consumption (approximate number of 8 oz. cups/glasses of:

a. Coffee: ____

b. Tea: ____

c. Caffeinated soda: ____

d. Other: ____

e. When was last time you had a caffeinated beverage? _______________________

APPENDIX C

IB QUESTIONNAIRE

Instructions: Please answer the following questions about the flight experiment.

1. What was your starting speed? ____________
2. What was your assigned speed change? ____________
3. What was your assigned altitude for the speed change? ____________
4. Approximately what altitude did you change speeds? ____________
5. What was your target landing altitude? ____________
6. Approximately what altitude did the scenario end at? ____________
7. Before the scenario ended, did you plan to land the aircraft? □ N □ Y
   Why or why not?

____

____

8. Did you notice anything unusual about the last few seconds of the landing? □ N □ Y
   If yes, please describe:

____

____

9. Did you notice anything unusual about the landing runway? □ N □ Y
   If yes, please describe:

____

____

10. Did you notice anything on or above the landing runway? □ N □ Y
    If yes, please describe:

____

____

11. Did you see an oil tanker? □ N □ Y
12. Did you see a stadium? □ N □ Y
13. Did you see a cellphone towers? □ N □ Y
14. Did you see a bus? □ N □ Y
    a. If Yes, please describe (e.g. color)____
15. Did you see a plane? □ N □ Y
    a. If Yes, please describe____
16. Did you see a tractor? □ N □ Y
    a. If Yes, please describe____
17. Did you see any other air traffic? □ N □ Y
    a. If Yes, how many other planes did you see? ____________________________________

APPENDIX D

NASA TLX WORKLOAD RATING SCALE

Workload Rating Scale
Based on the NASA Task Load Index (TLX) method used to assess workload on six scales.

Mental Demand: How mentally demanding was the task?

Low High

Physical Demand: How physically demanding was the task?

Low High

Temporal Demand: How hurried or rushed was the pace of the task?

Low High

Performance: How successful were you in accomplishing what you were asked to do?

Good Poor

Effort: How hard did you have to work to accomplish your level of performance?

Low High

Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Low High

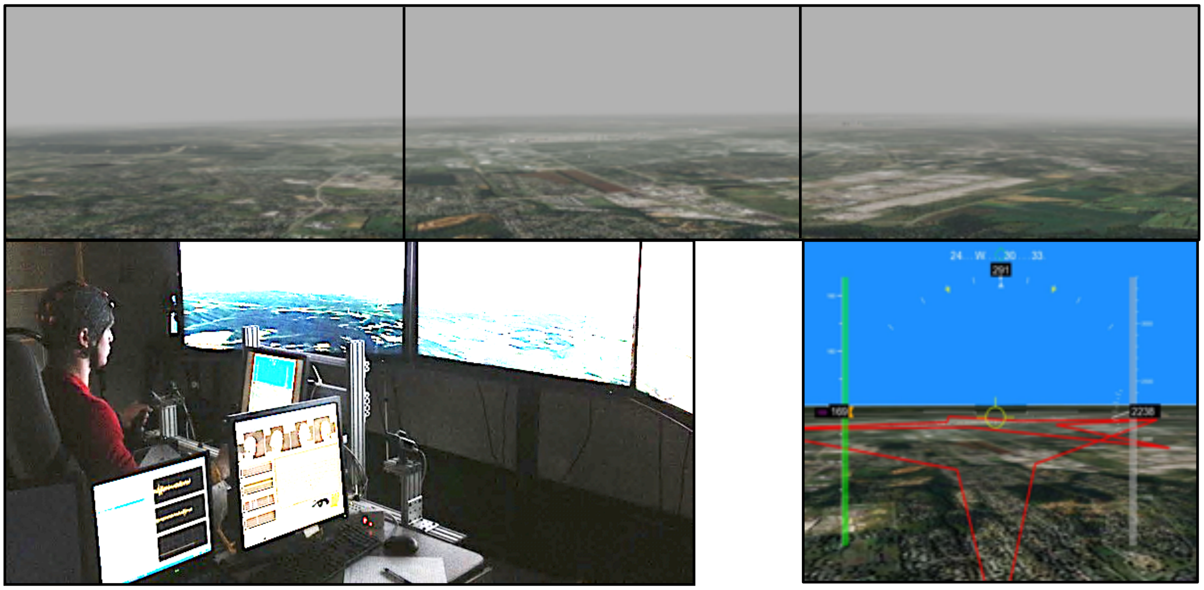
APPENDIX E

EXPERIMENT PROTOCOL

- 15 minutes: Introduction to experiment and subject informed consent.
- 30 minutes: Connecting subject to cortical/physiological monitoring equipment. Calibration of cortical/physiological monitoring equipment and SmartEye.
- 30 minutes: Completion of pre-experimental questionnaires (Appendix A and B).
- 5 minutes: Break
- 30 minutes: Training
    - Subjects will experience one of three automation conditions (full, partial, none) while landing a simulated aircraft in 3 pre-experimental training runs to familiarize subject with landing the simulated aircraft. More training is acceptable to ensure training to proficiency.
    - 10 min: After the final training run, subjects will complete a post-training questionnaire to indicate understanding of the speed and altitude changes required.
- 20 minutes: Testing
    - 10 min: Subjects will remain in the same automation condition as in training (full, partial, none) while landing a simulated aircraft at the same airport with very similar speed and altitude changes as training (simulation takes exactly 5 minutes).
    - 10 min: Subjects will complete the IB questionnaire (Appendix C) following Part 1.
    - 10 min: Subjects will complete the NASA-TLX (Appendix D) following Part 1.
- 10 minutes: Subject will be debriefed and disconnected from the cortical/physiological monitoring equipment.

APPENDIX F

PICTURE OF EXPERIMENTAL SETUP



Experimental set-up depicting Out-the-Window and Primary Flight Display visuals during a speed change.

APPENDIX H

RECRUITMENT FLYER

<u>ACTIVE STUDY SEEKING PARTICIPANTS!!!</u>

"The Impact of Automation on Operator Performance in a Simulated Landing Scenario"

This experiment is being conducted in the Human and Autonomous Vehicle Systems (HAVS) Laboratory at NASA Langley Research Center in Hampton. This laboratory uses a fix-based flight simulator and an array of physiological measurement tools to conduct experimental explorations.

The primary objective of this study is to compare non-pilot performance in a landing scenario across three automation settings (full manual, autothrottle, and autopilot). The secondary purpose of this experiment is to investigate the brain-generated electrical responses to virtual and non-virtual visual stimuli measured by electroencephalogram (EEG). This research will assist in the development of improved single pilot operations procedures.

Participants must be **non-pilots**, **aged 20 or older**, and have **no sensitivity to strobing light**. The experimental session will require approximately 3 hours to complete. Participants will be compensated $50.00. Civil Servant volunteers may participate in their official capacity.

Timeslots available from March 19 to March 28

APPENDIX I

PRIVACY ACT NOTICE

COLLECTION OF INFORMATION TO DETERMINE ELIGIBILITY TO PARTICIPATE IN RESEARCH AS A SUBJECT VOLUNTEER

GENERAL

This information is provided pursuant to Public Law 93-579 (Privacy Act of 1974), December 31, 1974, for individuals supplying information for inclusion in a system of records.

AUTHORITY

The authority to collect the information requested from you in the informed consent associated with **A study of inattentional blindness and steady state visually evoked potentials using cortical/physiological and self-report measures while performing a simulated flight task across different levels of automation** in which you may participate is derived from one or more of the following: Title 14, Code of Federal Regulations, Sections 1212 and 1230; Title 42, United States Code, Section 2451, as amended.

PURPOSES AND USES

The information you supply will is necessary to obtain your consent to participate in this research and to determine your eligibility to participate as a volunteer subject in the **A study of inattentional blindness and steady state visually evoked potentials using cortical/physiological and self-report measures while performing a simulated flight task across different levels of automation**. The information you provide will be evaluated by NASA employees and contractors overseeing and conducting the research. Your personal identifying information will not be shared outside of NASA and contractor and intern researchers working with NASA who are associated with this particular research. Your personal identifying information will be maintained under secure conditions (locked file), and only the Principal Investigator(s) (PI) overseeing your research will have access to your personal identifying information contained within the file.

The information will be maintained in a NASA System of Records: Human Experimental Research Data Records (NASA 10HERD). The information supplied is confidential and will be maintained under secure conditions as described above but is subject to routine uses for such information that are identified in System of Record Notice for Human Experimental Research Data Records published at 72 Federal Register 55812 on October 1, 2007. Release of such information is not permissible where your consent is required.

EFFECTS OF NONDISCLOSURE

Disclosure of the personal identifying information sought is voluntary; however, failure to furnish the information could exclude you from being able to participate as a volunteer in the research.

______________________________ Signature of Interviewer

______________________________ Signature of Volunteer

Date: ______________________________

APPENDIX J

EYE TRACKER SET UP AND PROCEDURE

All information reported herein pertains to Smart Eye Pro version 5.0 used for data collection as other versions have different or updated processes. The experimental system consisted of four 60 Hz Sony HR-50 6.0mm lens cameras (see Figure 10a at the end of the section) connected by ethernet, two IR-diode flashers, a 3-dimensional virtual representation of the critical portions of the testing environment (world model), and a 3-dimensional virtual representation of the subject's head (head model). The SE uses infrared diodes to provide consistent lighting across the subject's face and to produce glints, or cornea reflections of the IR flashers, that the system uses to locate the eye center. Multiple studies demonstrated successful pupillometry conducted in real-world driving applications with normally occurring lighting variations and free head movement (Recarte & Nunes, 2000, 2003). Smart Eye (2008) documentation states that this method improves gaze direction accuracy and reduces sensitivity to errors in head pose estimation as compared to eye center detection using head modeling alone. Glint detection reduces the gaze direction error related glint distortion due to ocular globe curvature changes and head pose estimation inaccuracies caused by large head movements or distorted facial expressions. SE stated these flashers were in compliance with the international standard IEC 62471 for "Photo-biological safety of lamps and lamp systems" and listed as having a 300-fold safety margin relative to the Maximum Permissible Exposure (MPE).

The Smart Eye system utilizes a 3D world model along with the 3D head model to detect gaze intersections with objects in the environment. The eye is an orb with a round pupil. An oblique viewing angle such as those garnered from peripheral cameras, can make a round pupil appear to be an oval and record a smaller diameter than the pupil would read if faced head-on. A system with an accurate head model with gaze direction awareness can help mitigate this error.

The researcher created 3D model representations of the four visual planes (3 out-the-window, 1 primary flight display) in the experimental environment at mm accuracy. The simulator utilized a four-camera system the width of the environment of interest and provide greater overlap of facial features viewed in each camera. The location of each camera and flasher was optimized by view of the subject eye and facial features in various head orientations and recorded in the Smart Eye environment demonstrating the location of the four 6.0mm cameras and two IR flashers placed on the frame of the simulator. From the subject position, all areas of interest were contained within a visual angle of 40º.

Positional requirements for the camera included remaining below the dominant gaze position, each eye visible in two or more cameras at any given time and proximity to each eye sufficient to produce a high-resolution image. Positional requirements for IR flashers included ensuring even lighting on the face and reduction of shadows. Pupillometry data required the flashers positioned least 10 centimeters from the optical axis. An example of this positioning is shown in Figure 10b. During operation, the simulator seat was fixed to maintain a constant distance of 26 inches from the center of the primary flight display screen defined as the point of origin for the eye tracking system.

Eye Tracking Procedure. First, the cameras were calibrated to each other using the predefined calibration process. The seat height was adjusted to position the subject into the appropriate vertical location by aligning the nose tip in the center of the Smart Eye software virtual head box. Each subject required a personalized 3D head model. First, the researcher would assess any need for focus and aperture adjustment for each Smart Eye camera as determined by continuous bars (see Figure 10c). Achieving optimal brightness was determined by maximizing a continual gray value histogram indicator on the display. This modification was rarely required as the room lighting and camera positions were fixed.

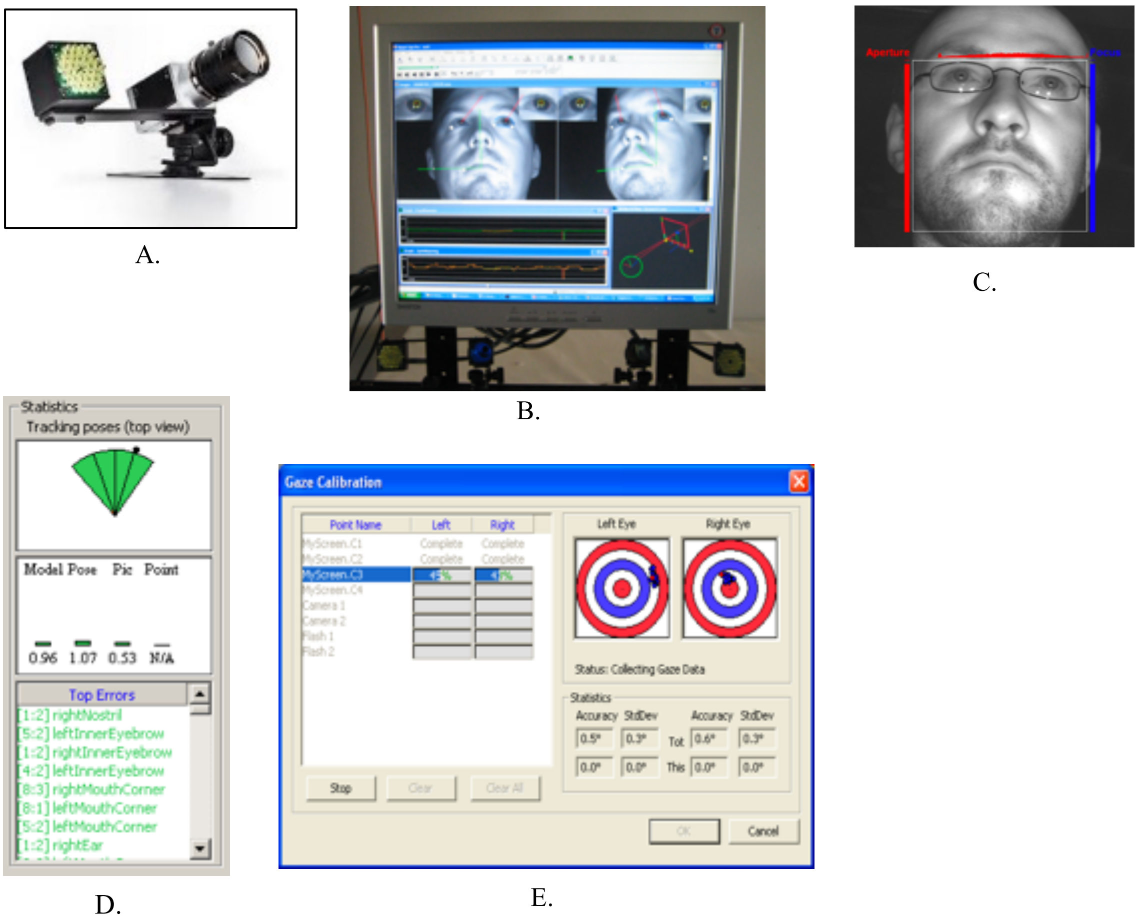
Next, the researcher captured a series of “snapshots” or images captured simultaneously across all four cameras with the subject oriented in specific gaze positions. These positions were representative across the expected gaze environment and worked from left to right across out-the-window and then repeated for primary flight display: 40º and 20º horizontally left, center, then 20º and 40º right of center. The subject looked straight ahead with the same gaze and head direction. Two additional positions were used to define the iris center and were captured with the subject gaze direction in the center of a camera while the head direction was oriented towards the center of the primary flight display.

Next, the researcher created a 3D head model by manually tagging prominent eye, ear, nose, and mouth features in each snapshot by pose and camera. Then, each head model was assessed overall and by pose and marker for accuracy in pixels (see Figure 10d). Error less than 3.0 pixels were color green and considered acceptable, orange was 3.0 - 5.0, and red was greater than 5. All profiles were corrected to produce a full green profile.

Finally, the head model was calibrated. Subjects oriented to fixation points placed on the out-the-window screens and the primary flight display co-located in the physical environment and virtual environment. The deviation between the physical orientation and virtual orientation was assessed with less than 2.0º accuracy error for each subject (see Figure 10e).

Figure 15

Examples of the Smart Eye setup and calibration procedure steps.



Note. From "Smart Eye Pro 5.0 - User Manual," By Smart Eye (2008). A. A smart eye camera with flasher. B. Smart Eye example of a two-camera tracking setup depicting eye feature marking with 3D head and world model. C. A Smart Eye example of the aperture and focus continuous scale and saturated pixel histogram guidance. D. A Smart Eye example of error information for each pose and feature marked in that pose.

VITA

Kellie D. Kennedy, M.S.
Department of Psychology
250 Mills Godwin Life Sciences Building
Norfolk, VA 23529

Education

2021 Old Dominion University, Norfolk, VA
Ph.D., Psychology - Human Factors
Advisor: Michelle L. Kelley, Ph.D.

2015 Old Dominion University, Norfolk, VA
Completed Candidacy Exams
Major Area: Human Factors Psychology
Area of Concentration: Visual attention failures in the flight deck.

2012 Old Dominion University, Norfolk, VA
M.S., Psychology - Applied Experimental
Advisor: James Bliss, Ph.D.
Focus: Industrial/Organizational & Human Factors Psychology
Thesis: Inattentional blindness in a simulated navigational driving task.

2011 Old Dominion University, Norfolk, VA
Modeling and Simulation (M&S) Graduate Certificate

2009 Wright State University, Dayton, Ohio
B.S., Experimental Psychology, Summa Cum Laude, Departmental Honors
Advisor: Nathan Bowling, Ph.D.
Focus: Industrial/Organizational & Human Factors Psychology
Minor: Management
Honors Thesis: The effects of perceived organizational support on organizational citizenship behaviors as moderated by conscientiousness.

Employment

May 2013 – present NASA Langley Research Center
Crew Systems and Aviation Operations Research Branch
Research Aerospace Engineer – Human Factors

Notable Award

2018 - Awarded the NASA Exception Achievement Medal for work as sole Human Factors Subject Matter Expert on the NASA Engineering and Safety Center (NESC) special team conducting a congressionally mandated investigation regarding physiological events occurring in pilots flying the F/A-18 jet aircraft with the U.S. Navy.

Patents

LAR-19816-1 - Titled: Display System Interface Using Visually-Evoked Cortical Potentials. Initiated patent process on 5-29-2020.

LAR-19051-1-CON - Titled: System and Method for Human Operator and Machine Integration. Awarded 05-07-2021.