

From the lab to the cockpit: the development and analysis of quantifiable measures of goal-directed actions through gaze behaviour

by
Naila Ayala

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Kinesiology

Waterloo, Ontario, Canada, 2024

© Naila Ayala 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisor	Dr. Ewa Niechwiej-Szwedo Associate Professor, Faculty of Health, Department of Kinesiology and Health Sciences, University of Waterloo
Thesis Advisory Committee	Dr. Sean Meehan Associate Professor, Faculty of Health, Department of Kinesiology and Health Sciences, University of Waterloo Dr. Elizabeth Irving Professor, Faculty of Science, School of Optometry & Vision Science, University of Waterloo Dr. Suzanne Kearns Associate Professor, Faculty of Environment, Department of Geography & Environmental Management, University of Waterloo
Internal- External Examiner	Dr. Evan Risko Associate Professor, Faculty of Psychology, University of Waterloo
External Examiner	Dr. Luc Tremblay Professor, Faculty of Kinesiology and Physical Education, University of Toronto

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I declare that I was the main author for all writing of the thesis and main contributor of the research design, data collection, data analysis and data interpretation, which were performed under the supervision of Dr. Ewa Niechwiej-Szwedo. Dr. Ewa Niechwiej-Szwedo was the primary co-author on all published manuscripts which were derived from this thesis. Dr. Elizabeth Irving, Dr. Shi Cao, and Dr. Suzanne Kearns were all secondary co-authors on studies two to five in this manuscript. Abdullah Zafar was another co-author on studies one, two and four in this manuscript and generated the scripts used for all entropy and bout analyses. Dr. Diako Mardanbegi and Dr. Andrew Duchowski were co-authors on study four in this manuscript.

Abstract

Objectives: The goal of this thesis was to examine the utility of gaze behaviour as an indicator of information processing during the performance of complex tasks in ecological environments. Information processing is influenced by the interaction between task requirements, an individual's resource availability, and the environmental context (i.e., modifying factors). Five experiments were performed that manipulated these modifying factors and examined the respective changes across low fidelity and high-fidelity environments. Study one comprehensively examined gaze behaviour while participants performed a neurocognitive test where task demands increased progressively by loading the working memory system (i.e., cognitive load). Studies two and three characterized gaze behaviours associated with increased sensorimotor task demands during the performance of simulated landing scenarios using a computer-based simulation and high-fidelity flight simulator, respectively. Study four aimed to explicitly assess differences in gaze behaviour between novices (i.e., ab-initio) and experts (i.e., licensed low-time pilots) during a simulated flight task, while study five expanded on findings from these studies by examining gaze behaviour during an extended (i.e., cruise and landing) flying scenario that involved a dual-task paradigm (i.e., cognitive load).

Methods: Twenty-seven participants were recruited for study one. Task performance and eye movements were recorded as they completed the Tower of London task, which progressively increased in cognitive load by manipulating problem set size (i.e., minimum number of moves to solve the task). In studies two and three, sensorimotor difficulty was manipulated by altering the wind conditions during a landing scenario (i.e., no wind, turbulent winds). Eighteen participants in study two and twenty-nine participants in study three were recruited and had their gaze behaviour recorded during the task in a low-fidelity (i.e., PC-based) and high-fidelity (i.e., fully immersive flight training device) simulation environment, respectively. Study four examined gaze behaviour and performance differences during a simulated landing task between ab-initio pilots (i.e., 0 flight hours) and licensed low-time pilots (i.e., on average 193 flight hours). Study five examined secondary auditory oddball performance, gaze behaviour and flight performance as a function of task condition (single, dual) and flight phase (cruise, landing).

Results: Results from the first four studies revealed several uses for traditional (i.e., dwell time, fixation rate, fixation duration, scan path length) and advanced (i.e., SGE, GTE, cognitive tunneling analysis) gaze behaviour metrics. Specifically, when task requirements increased (i.e., sensorimotor demands, cognitive load) the focus of attention became more biased toward areas that held more task-critical information and were necessary for successful task performance. Many of these measures were also sensitive to phase specific changes in gaze behaviour associated with distinct task goals. Indeed, overall shifts in gaze dispersion, dwell time, and scanning patterns were shown to reflect sources of information that were important for achieving flight task specific sub-goals. Such differences were evident independent of changes in sensorimotor difficulty or cognitive load. These changes in gaze behaviour due to variation in task requirements were consistent across low-fidelity and high-fidelity cockpit environments. Several gaze behaviour measures also demonstrated that experienced pilots used vision more efficiently and effectively, by sampling more task-related regions of the environment more frequently with fewer instances of cognitive tunneling. The last aviation study was the only paradigm that did not show significant differences in gaze behaviour or flight performance due to increased cognitive load. This was despite the presence of increased auditory response times and

error rates. Although most gaze metrics examined here were modulated (or influenced) by a number of factors impacting skill performance, blink rate seemed to be the one measure that was not consistent.

Conclusions: Gaze behaviour analysis provides important spatiotemporal information related to the underlying cognitive processes supporting complex skill performance. Gaze behaviour metrics are sensitive indicators of various task requirements and individual resource availability. It is important to note that the gaze behaviours reported are highly task dependent and are stable across the environmental contexts examined here. Collectively, the gaze metrics examined here provide valuable information for assessing information processing and can contribute to better characterization of visual scanning in complex environments.

Acknowledgements

Life shrinks or expands in proportion to one's courage. – Anais Nin

If the last five years of my life can attest to anything, it is that courage can indeed expand one's life beyond their wildest dreams. During my time at the University of Waterloo I have been fortunate enough to receive unparalleled access to incredible resources and world-class experts in science and aviation. However, the quality and quantity of work I have been able to produce over the years is primarily a product of the community that has helped build and nurture the fire within me that is my imagination, my curiosity, and most importantly, my courage to do things that have not been done before. For this, I have many people to thank.

First and foremost, I would like to acknowledge and thank Dr. Elizabeth Irving, Dr. Suzanne Kearns, Dr. Shi Cao, and Dr. Ewa Niechwiej-Szwedo who have all guided, supported and challenged me to become the scientist I am today. Thank you for welcoming me into WISA as the first PhD student and providing me with endless opportunities to learn and develop as a scientist with one foot in academic research and the other in industry. In this environment, I have found my voice, my motivation, and my passion for research that has truly made this scientific journey an adventure of a lifetime that I hope will never end. Moreover, I am endlessly grateful to Dr. Ewa Niechwiej-Szwedo for welcoming me into the Developmental Visuomotor Neuroscience Lab and providing me with an amazing training environment where I could pursue multiple projects, become inspired to develop novel research ideas, and more importantly, receive endless feedback on how to improve the caliber of my work and grow as a person (in academia and in life).

Second, I would like to thank my lab mates and friends. Your support and laughter have helped me get through the highs and lows over the past five years. What I will miss the most, is our aviation meetings where we spent hours of time looking at data, coming up with novel methods to analyze data, and having my eyes opened to the beauty of math.

To my family, the accomplishment of my PhD is as much a testament to my hard work, dedication, and perseverance as it is to all of yours. I would like to acknowledge and thank my parents for actively showing me what it means to give something your all and giving me the space and time to make my failures learning opportunities. Your sacrifice, love, and support know no limits and have allowed me to be courageous in my actions and in my aspirations. To all my siblings, your paths ahead are and will continue to be different from mine, but that doesn't make them any less difficult. Many of us have already shared in each others' tears, daunting late nights, and laughter, but the other thing that we will continue to share is this invisible string that connects us all despite our own respective journey's. Throughout my life, you have all held me close and allowed me to lean on you when I needed to. I am committed to providing that same level of connection and support to all of you whenever and wherever, as this has allowed me to stay committed to my goals and be my true self to a fault.

I would also like to acknowledge and thank my examining committee and external examiner. Your questions, insights, and feedback all provided for a thought-provoking defense, and I look forward to incorporating some of the discussed topics in future research.

Lastly, to the most courageous person I know, my wife. Thank you for being my circuit breaker, my sounding board, and my biggest motivator. I love you, endlessly.

Table of Contents

Examining Committee Membership.....	ii
Author's Declaration.....	iii
Statement of Contributions.....	iv
Abstract.....	v
Acknowledgements.....	vii
List of Figures.....	xi
List of Tables.....	xvii
List of Abbreviations.....	xix
 Section 1: General Introduction.....	 1
1.1 Gaze behaviour in complex task performance.....	1
 Section 2: Literature review.....	 5
2.1 Oculomotor and gaze control.....	5
2.1.1 The neural substrates involved in gaze behaviour.....	6
2.2 Overview of executive functions.....	8
2.3 Eye movements and their applications to the real world.....	10
 Section 3: Research aims, questions and hypotheses.....	 15
3.1 Research aims.....	15
3.2 Overview of studies: Research questions and hypotheses.....	18
 Section 4: <i>Gaze Behaviour: A window into distinct cognitive processes revealed by the Tower of London test</i>	 24
4.1 Introduction.....	24
4.2 Methods	30
4.2.1 Participants.....	30
4.2.2 Materials and procedures.....	31
4.2.3 Data reduction.....	33
4.2.4 Performance analysis.....	34
4.2.5 Traditional gaze analysis.....	35
4.2.6 Entropy analysis.....	35
4.2.7 Statistical analysis.....	37
4.3 Results.....	38
4.3.1 Performance measures.....	38
4.3.2 Gaze behaviour during the planning interval.....	39
4.3.3 Gaze behaviour during the execution interval.....	43
4.3.4 The influence of task difficulty on gaze entropy.....	44
4.3.5 Comparison of optimal and non-optimal performance.....	46
4.4 Discussion.....	52
4.4.1 Performance indicators of increasing task difficulty and non-optimal performance....	53
4.4.2 Gaze characteristics during planning.....	54
4.4.3 Gaze characteristics during solution execution.....	57
4.4.4 Gaze characteristics distinguish optimal from non-optimal performance.....	59
4.4.5 Limitations.....	63
4.5 Conclusion.....	64
 Section 5: <i>The effect of task difficulty on gaze behaviour during landing with visual flight rules in low-time pilots</i>	 66
5.1 Introduction.....	66
5.2 Materials and methods	72
5.2.1 Participants.....	72
5.2.2 Apparatus.....	73
5.2.3 Scenario and task.....	73

5.2.4 Data reduction.....	76
5.2.5 Entropy analysis.....	77
5.3 Results.....	78
5.3.1 Performance measures.....	78
5.3.2 Traditional gaze measures.....	81
5.3.3 Entropy-based measures.....	85
5.4 Discussion.....	90
5.4.1 The effect of task difficulty on performance.....	91
5.4.2 The effect of task difficulty on gaze behaviour.....	91
5.4.3 Limitations.....	96
5.5 Conclusion.....	97
Section 6: <i>Investigating the role of flight phase and task difficulty on low-time pilot performance, gaze dynamics and subjective situation awareness during simulated flight</i>	
6.1 Introduction.....	99
6.2 Methods	104
6.2.1 Participants.....	104
6.2.2 Experimental setup and apparatus.....	105
6.2.3 Scenario and task.....	106
6.2.4 Data reduction.....	109
6.2.5 Performance and gaze analysis.....	110
6.2.6 Statistical analysis.....	114
6.3 Results.....	115
6.3.1 The effect of task difficulty on landing performance and aircraft control.....	115
6.3.2 The effects of task difficulty on gaze behaviour.....	119
6.3.3 The effects of task difficulty on situation awareness.....	128
6.3.4 Regression analyses.....	129
6.4 Discussion.....	131
6.4.1 Examining the effectiveness of the task difficulty manipulation.....	132
6.4.2 Flight phase parameters and experience influence information processing.....	139
6.4.3 Limitations and future directions.....	143
6.5 Conclusion.....	144
Section 7: <i>Does fiducial marker visibility impact task performance and information processing in novice and low-time pilots?</i>	
7.1 Introduction & background.....	146
7.1.1 Paper overview and contributions.....	147
7.1.2 Related work.....	149
7.2 Invisible markers.....	153
7.3 Empirical validation.....	154
7.3.1 Experimental procedure.....	155
7.3.2 Participants.....	157
7.3.3 Experimental setup and apparatus.....	158
7.3.4 Data processing and analysis.....	159
7.4 Results.....	157
7.5 Discussion.....	168
7.6 Limitations and future work.....	172
7.7 Conclusion.....	174
Section 8: <i>The effects of a dual task on gaze behaviour examined during a simulated flight in low-time pilots</i>	
8.1 Introduction.....	176
8.2 Methods.....	180
8.2.1 Participants.....	180
8.2.2 Experimental setup and apparatus.....	181
8.2.3 Scenario and task.....	183

8.2.4 Data processing and analysis.....	186
8.3 Results.....	191
8.3.1 Subjective Situation Awareness Ratings.....	191
8.3.2 Auditory task performance.....	192
8.3.3 Flight task performance.....	194
8.3.4 Gaze behaviour.....	195
8.4 Discussion.....	199
8.5 Limitations and future directions.....	205
8.6 Conclusion.....	206
Section 9: <i>Summary of Research</i>	207
9.1 General discussion.....	207
9.2 Limitations and future directions.....	210
9.3 Conclusion.....	215
References.....	217

List of Figures

3.1 – An overview of information processing and the contributing factors underlying skill performance. Gaze behaviour refers to eye movements which serve to acquire visual information. Since the area of high resolution (i.e., spatial acuity) is limited, examining gaze behaviour highlights where and when attention is being deployed based on environmental factors (i.e., simulated/real environments, presence of evaluators), individual resource availability (i.e., level of experience), and task requirements (i.e., sensorimotor demands and cognitive load). The manner through which gaze is selectively deployed is driven by competing bottom-up and top-down processes that are governed by Executive Functions (EFs). Understanding the appropriate balance of top-down and bottom-up processes to meet the needs of the environment, individual, and task has critical implications for skill performance and learning.	17
4.1 – Illustration of the visual stimuli employed in the Freiburg Version of the Tower of London task. Goal space is displayed in the top array and workspace is displayed in the bottom array. The minimum number of moves required for the problem is indicated to the left of the workspace. The blue boxes represent the two main areas of interest used in the traditional gaze analysis. The orange boxes represent the six main areas of interest used in the entropy analysis	32
4.2 - Tower of London performance measures. Group mean accuracy scores (%) (A), initial planning time (sec) (B), and movement execution time (sec) (C) are demonstrated across all levels of task difficulty. Only correct trials are reflected in panels (B) and (C). Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$	39
4.3 - Group mean dwell time (%) during the planning interval (A) and execution (B) and scan path length (°) during the planning interval (C) and execution (D) are demonstrated across all levels of task difficulty. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$	41
4.4 - Group mean initial planning interval (sec) when participants looked in the goal space versus the workspace first. Error bars represent SEM	42
4.5 - Group mean entropy (bits) values for stationary gaze entropy (SGE) (A and B) and gaze transition entropy (GTE) (C and D) during the planning and execution intervals, respectively, across all levels of task difficulty. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$	45
4.6 - Group mean initial plan time (sec) (A) and movement execution time (sec) (B) are illustrated for optimal (black) and non-optimal (grey) data as a function of increasing task difficulty. Error bars represent SEM. *** $p \leq 0.001$	47
4.7 - Group mean fixation rate (fixations per sec) (A) and fixation duration (ms) (B) during the planning interval are illustrated for optimal (black) and non-optimal (grey) data. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$	48

4.8 - Group mean scan path length (°) is demonstrated across all levels of task difficulty for optimal (A) and non-optimal (B) responses during the execution interval. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$	49
4.9 - Group mean entropy (bits) values are shown for optimal (filled black circles) and non-optimal (open circles) stationary gaze entropy (SGE) (A and B) and gaze transition entropy (C and D) across task difficulty levels 4-6 during the planning and execution intervals, respectively. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$	51
5.1 - Illustration of the visual stimuli employed in the Microsoft Flight Simulator landing challenge environment. The participants point of view of the cockpit replicated that of a pilot flying a Cessna 152, pre-set for a straight-on approach to Billy Bishop airport, Toronto, Canada. The orange boxes represent the ten main areas of interest used in the gaze analyses. These include the airspeed (1), attitude (2), altimeter (3), turn coordinator (4), heading (5), vertical speed (6) and power (7) indicators, as well as the runway (8), horizon (9), and side window (10).....	75
5.2 - Individual data points and their respective group means for success rate (%) (A), completion time (sec) (B), and overall performance scores (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$	80
5.3 – Individual data points and their respective group means for landing smoothness (feet per minute:fpm) (A), landing accuracy (ft) (B), and ground roll (ft) (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$	81
5.4 – Illustration of the visual stimuli employed during landing and the corresponding change in dwell time (%) between easy and difficult conditions. Only significant changes in dwell time are displayed over their corresponding area of interest (AOI) (i.e., airspeed, power, runway, and horizon AOIs). Group mean differences (difficult-easy) are illustrated in red to indicate a decrease in AOI dwell time and green to indicate an increase in AOI dwell time	84
5.5 - Individual data points for stationary gaze entropy (SGE) (bits) (A) and gaze transition entropy (GTE) (B) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$	86
5.6 - Representative time traces that reflect the stationary gaze entropy (SGE) and gaze transition entropy (GTE) of a single easy (A) and difficult (B) trial. Exemplar time traces showing the probability of fixating inside the cockpit for the easy (C) and difficult (D) conditions. These reflect the same easy (A and C) and difficult (B and D) trials. Gray boxes represent the sections of the trial where entropy and P(inside) equals zero. This indicates that gaze is directed outside the cockpit for at least 10 seconds (i.e., a single bout). The easy condition illustrates four bouts with a total bout time of 68 sec (A). The difficult condition illustrates six bouts with a total bout time of 96 sec (B).....	88

5.7 - Individual data points for number of bouts (A), bout duration (sec) (B), and total bout time (sec) (C) are demonstrated for easy and difficult conditions. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$ 90

6.1 - Illustration of the simulator and eye tracking equipment set-up. The ALSIM simulator was set up as a single-engine aircraft, controlled with the yolk, throttle lever, and rudders pedals (A). Participants also used a headset (hanging on the left window) to make ATC calls (A). The AdHawk glasses (B) were worn by participants throughout the experimental session and were connected to the collection laptop (B) via a USB-C cord. The collection laptop used AdHawk software to calibrate the eye tracker and start and stop eye tracking data collection 106

6.2 - Illustration of the flight path and its respective phases of flight encountered in the experimental landing scenarios: Downwind, Base, Final Approach 107

6.3 - Illustration of the visual stimuli employed in the AL250 flight simulator environment. The participants point of view of the cockpit replicated that of a pilot flying a Cessna 172, pre-set for a downwind-to-base-to-final approach to Waterloo International Airport, Breslau, Ontario, CA. The orange boxes represent the ten main areas of interest used in the gaze analyses. These include the airspeed (1), attitude (2), altimeter (3), VOR (4), heading (5), vertical speed (6) and power (7) indicators, as well as the front (8), left (9), and right windows (10) 110

6.4 - Individual data points and their respective group means for completion time (A), landing hardness (B), and landing accuracy (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 116

6.5 - Individual data points and their respective group means for average airspeed RMSE are demonstrated for easy and difficult conditions across all downwind, base, and final approach stages of flight. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 118

6.6 - Individual data points and their respective group means for vertical speed (VSpeed) (kts) (A), and VSpeed root mean square error (RMSE) (B) are demonstrated for easy and difficult conditions across the final approach stage of flight. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 119

6.7 - Candy-bar plots illustrate the overall dwell time (%) allotments across all AOIs during their respective phases of flight (A, B, C) for easy (left panels) and difficult (right panels) conditions 121

6.8 - Individual data points and their respective group means for blink rate (blinks/sec) are demonstrated for each phase of flight. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 125

6.9 - Normalized individual data points and their respective group means for stationary gaze entropy (SGE) (bits) (A) and gaze transition entropy (GTE) (B) are demonstrated for easy and

<i>difficult conditions each phase of flight (i.e., downwind, base, final approach) . Error bars represent SEM. *p≤0.05, **p≤0.01, ***p≤0.001</i>	<i>127</i>
<i>6.10 - Individual data points and their respective group means for number of cognitive tunneling bouts (A) average bout duration (B), and total cognitive tunneling bout time (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. *p≤0.05, **p≤0.01, ***p≤0.001</i>	<i>128</i>
<i>6.11 - Individual data points and their respective group means for Situation Awareness (SA) Score (A), SA Understanding subcomponent score (B), SA Supply subcomponent score (C), and SA Demand subcomponent score (D) are demonstrated for easy and difficult conditions. Error bars represent SEM. *p≤0.05, **p≤0.01, ***p≤0.001</i>	<i>129</i>
<i>6.12 - Scatter plots demonstrate the relationship between the number of cognitive tunneling bouts and pilot flight hours (A), and the total cognitive tunneling bout time (sec) and pilot flight hours (B). Black circles indicate easy condition participant averages, red triangles indicate difficult condition participant averages. Lines indicate the trend for all subjects (i.e., line of best fit). Highlighted bands around line of best fit indicate the 95% confidence interval</i>	<i>130</i>
<i>7.1 - Illustration of the ArUco marker placement within the flight simulator for both visible (left) and invisible (right) marker types</i>	<i>153</i>
<i>7.2 - IR longpass filter marker covering (left) and IR light source used in the experiment (right). The front side of IR light source was directed toward the markers inside the simulator.....</i>	<i>154</i>
<i>7.3 – Examples of the camera view during the experiment: (left) image of the cockpit with original ArUco markers with a close-up of a marker, and (right) image of an invisible ArUco marker with close-up of the marker. Notice that the markers are visible to the camera in both instances (in right image due to IR illumination) although invisible markers appear slightly dimmer</i>	<i>155</i>
<i>7.4 – Illustration of the visual stimuli employed in the simulator environment. The participant’s point of view in the cockpit was pre-set for a downwind-to-base-to-final approach to Waterloo International airport, Breslau, ON, Canada. The orange boxes represent the fourteen Areas Of Interest used in the gaze analyses. These include the (1) airspeed, (2) attitude, (3) altitude, (4) VHF Omni Range (VOR), (5) heading, (6) vertical speed, (7) power, as well as (8) the front window, (9) left window, (10) right window. The additional 4 Areas of Interest (11-14) represent the four markers that were placed in the cockpit around the instrument panel for optimal camera detection.....</i>	<i>157</i>
<i>7.5 – AdHawk Mindlink eye tracker used in the study</i>	<i>159</i>
<i>7.6 – Individual participant data points and their respective group means for (left) completion time expressed in seconds, (center) landing accuracy expressed in degrees, and (right) landing</i>	

<i>hardness expressed in feet per minute (fpm). Error bars represent SEM (*p<0.05, **p<0.01).....</i>	<i>165</i>
<i>7.7 – Individual participant data points and their respective group means for (left) Stationary Gaze Entropy (SGE), expressed in bits, (center) Gaze Transition Entropy (GTE), expressed in bits, and (right) fixation frequency (i.e., count). Error bars represent SEM (*p<0.05, **p<0.01).....</i>	<i>166</i>
<i>7.8 – Individual participant data points and their respective group means for (left) number of gaze tunneling bouts (i.e., count), and (right) total gaze tunneling bout time (sec). Error bars represent SEM (*p<0.05, **p<0.01).....</i>	<i>167</i>
<i>7.9 – Individual participant data points and their respective group means for subjective distraction scores. Error bars represent SEM (*p<0.05, **p<0.01).....</i>	<i>168</i>
<i>8.1 - Illustration of the visual stimuli employed in the AL250 flight simulator environment. The participants point of view of the cockpit replicated that of a pilot flying a Cessna 172, pre-set for a cruise and landing approach to Waterloo International Airport, Breslau, Ontario, CA. The orange boxes represent the ten main areas of interest used in the gaze analyses. These include the airspeed (1), attitude (2), altimeter (3), turn coordinator (4), heading (5), vertical speed (6) and power (7) indicators, as well as the front (8), left (9), and right windows (10).....</i>	<i>182</i>
<i>8.2 - Illustration of the simulated flight path and circuit parameters for cruise (easy [blue]) and landing (difficult[pink]) phases of flight.....</i>	<i>185</i>
<i>8.3 - Individual data points and their respective group means for situation awareness (SA) (A), SA understanding (B), SA supply (C), and SA demand are demonstrated for single and dual task conditions. Error bars represent SEM. *p≤0.05, **p≤0.01, ***p≤0.001.....</i>	<i>192</i>
<i>8.4 - Individual data points and their respective group means for auditory response time cost (%) (A) and response error cost (%) are demonstrated for cruise and landing phases of flight. Error bars represent SEM. *p≤0.05, **p≤0.01, ***p≤0.001.....</i>	<i>193</i>
<i>8.5 - Group means for dwell time (%) (A) and dwell duration (msec) (B) differences (Δ: dual-single) across all areas of interest (AOIs) during cruise (open circle) and landing (open triangle) phases of flight. Error bars represent SEM. Significant interactions between task and flight phase are indicated accordingly *p≤0.05, **p≤0.01, ***p≤0.001.....</i>	<i>196</i>
<i>8.6 - Individual data points and their respective group means for stationary gaze entropy (SGE) (A), gaze transition entropy (GTE) (B), number of cognitive tunneling bouts (C), and total cognitive tunneling bout time (sec) (D) are demonstrated for both task conditions (single, dual)</i>	

across cruise and landing phases of flight. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$,
 *** $p \leq 0.001$ 198

List of Tables

4.1 – Normalized group mean entropy values for stationary gaze entropy (SGE) and gaze transition entropy (GTE) during the planning and execution intervals, respectively, across all levels of task difficulty. Standard deviation values reported in brackets 46

4.2 – Normalized group mean entropy values for non-optimal stationary gaze entropy (SGE) and gaze transition entropy (GTE) during the planning and execution intervals, respectively, across task difficulty levels 4-6. Standard deviation values reported in brackets 52

5.1 – Traditional gaze values calculated for all areas of interest during easy and difficult conditions.

Note: Mean (standard deviation) values for all traditional gaze measures across all areas of interest (AOI) and task conditions (easy versus difficult). Significant changes between task difficulties and their corresponding dependent variable and AOI reported via * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$ 82

6.1- Dwell time (%) values calculated for all areas of interest for easy and difficult conditions across all phases of flight.

Note: Mean (standard deviation) dwell time (%) values across all areas of interest and task difficulty levels (i.e., easy, difficult). Significant changes between easy and difficult conditions are reported via * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 122

6.2- Dwell rate (dwells/sec) values calculated for all areas of interest for easy and difficult conditions across all phases of flight.

Note: Mean (standard deviation) dwell rate (dwells/sec) values across all areas of interest and task difficulty levels (i.e., easy, difficult). Significant changes between easy and difficult conditions are reported via * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 123

6.3- Average dwell duration (msec) values calculated for all areas of interest across all phases of flight.

Note: Mean (standard deviation) average dwell duration (msec) values across all areas of interest and phases of flight (i.e., downwind, base, final). Significant changes between downwind and base phase of flight are reported via # $p \leq 0.05$, ## $p \leq 0.01$, ### $p \leq 0.001$. Significant changes between base and final phase of flight are reported via ^ $p \leq 0.05$, ^^ $p \leq 0.01$, ^^ $p \leq 0.001$. Significant changes between downwind and final phase of flight are reported via * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 124

6.4- Entropy (bits) values calculated for all areas of interest for easy and difficult conditions across all phases of flight.

Note: Mean (standard deviation) entropy (bits) values across all phases of flight (i.e., downwind, base, final) and task difficulty levels (i.e., easy, difficult). Significant changes between easy and difficult conditions are reported via * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 127

6.5- *The associations between task difficulty and flight hours for number of bouts and total bout time.*

Note: Estimated marginal (EM) mean of linear trend characteristics for the number of bouts and total bout time models are provided including their respective coefficients (β), 95% lower/upper confidence levels, and R-squared values.....130

8.1- *List of dependent measures and their associated ANOVAs and independent variables*

Note: The listed dependent variables represent the larger clusters of actual dependent variables. Flight performance metrics include landing performance metrics (total completion time, landing error, landing hardness), as well as cruise performance metrics (airspeed and vertical speed mean and variability). Raw auditory performance metrics include response time and response accuracy. Dual task cost auditory scores include auditory response cost and auditory accuracy cost. Gaze behaviour metrics include dwell time, fixation duration, SGE, GTE, cognitive tunneling bout number, duration, and total time. SART scores include overall SA, SA supply, SA demand, and SA understanding scores.....191

8.2- *Flight performance values for single and dual task conditions.*

Note: Mean (standard deviation) flight performance values are reported for single and dual task conditions. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ 194

8.3- *Dwell time (%) descriptive and group statistics*

Note: Mean (standard deviation) dwell time (%) values across significant areas of interest (AOI) are reported across cruise and landing phases of flight, which were collapsed across single and dual task conditions. Statistical test values (F , p , η_p^2) for main effects of phase of flight (i.e., cruise, landing) are reported for only significant AOIs. Df (1,20).....196

8.4- *Average dwell duration (msec) descriptive and group statistics*

Note: Mean (standard deviation) average dwell duration (msec) values across significant areas of interest (AOI) are reported across cruise and landing phases of flight. Statistical test values (F , p , η_p^2) for main effects of phase of flight (i.e., cruise, landing) are reported for only significant AOIs. Df (1,20).....197

List of Abbreviations

AOI – Area's of Interest

CPL- Commercial Pilot's License

DLPFC – Dorsolateral prefrontal cortex

DTC- Dual Task Cost

EF – Executive function

FPM- Feet per minute

FEF – Frontal eye field

GTE – Gaze transition entropy

HMPD- Head mounted projected display

IFR- Instrument flight rules

LIP – Lateral intraparietal area

MT – Movement time

MT area – Middle temporal area (V5)

PD – Parkinson's disease

PPL – Private Pilot's License

PPRF – Pontine paramedian reticular formation

RT – Reaction time

SA – Situation awareness

SAGAT – Situationa awareness global
assessment technique

SART – Situational awareness rating technique

SPAM – Situational present assessment method

SC – Superior colliculus

SEF – Supplementary eye field

SGE – Stationary gaze entropy

TOL – Tower of London

TOL-F – Tower of London- Freiburg version

VFR – Visual flight rules

VOR – Vestibular ocular reflex

Section 1: General introduction

1.1 Gaze behaviour in complex task performance

The successful completion of everyday activities relies on complex interactions between visual, motor and cognitive processes (de Brouwer et al., 2021). Whether that be driving a car, reading a book, or making a cup of coffee, humans interact with the world around them by moving their head and eyes to focus their line of sight (i.e., gaze) toward a specific object or location of interest. Shifting gaze to redirect the line of sight serves two main goals. First, the photoreceptor distribution varies across the retina such that high resolution vision is confined to the central region referred to as the fovea (Kandel et al., 2013). Therefore, eye movements ensure that the fovea is aligned with objects or areas of interest (AOIs) to allow for detailed visual information processing. Second, the brain has a limited capacity to selectively process and manipulate information in working memory (Oberauer, 2019). Therefore, we must choose which objects to focus our attention on and which ones to ignore. The selection of objects is driven by the interaction between top-down (i.e., goal driven) and bottom-up (i.e., saliency driven) processes – these competitive interactions are influenced by executive functions (EFs) (Lavie et al., 2004). Notably, EFs are a set of higher-level cognitive processes underlying our ability to selectively attend to specific information (selective attention), inhibit prepotent responses in favor of alternative commands (inhibitory control), hold and manipulate information in working memory (working memory), and shift mental task/goal sets (cognitive flexibility). Collectively, EFs facilitate the selection, prioritization and processing of task-related information that support goal-directed behaviours (Diamond, 2013). As such, where we look is a good indication of what information we are overtly attending to (Gegenfurtner, 2016) – though this is not always the case, for example covert attention allows processing of information that is not fixated or

inattention blindness where fixated information is seemingly not processed (Gegenfurthner, 2016; Kandel et al., 2013). Nevertheless, eye movements, and thus, the selection and processing of visual information in our environment relies on EFs to manage the interplay between bottom-up and top-down processing of stimuli based on environmental context, task requirements, and the individual's resource availability. Extensive research over the past few decades has mapped out the associations between eye movements and many aspects of EFs such as inhibitory control/selective attention, working memory, and cognitive flexibility; however, these processes have been studied mostly in the context of highly controlled lab-based paradigms. A major gap in knowledge remains to systematically examine the interaction between eye movements and information processing during more complex tasks in ecological environments.

Eye movements have been used extensively as a non-invasive behavioural assay for examining EFs (Anderson et al., 2012; Coubard, 2013; Ford et al., 2005; Kaller et al., 2012; Kowler, 2011; Krauzlis, 2005; Norman and Shallice, 1986; Munoz and Everling, 2004). This body of knowledge is based on the exhaustive work done in both animal models and humans demonstrating that the neural circuits controlling EFs and eye movements are overlapping and closely linked (Wurtz and Albano, 1980; Neggers et al., 2005; Moore and Fallah, 2001; Munoz and Everling, 2004; Pierrot-Deseilligny et al., 2003; Royall et al., 2002). For instance, the pre-frontal circuits involved in the preparation of a saccade (i.e., eye movement) play a causal role in the direction of overt and covert visuospatial attention (Awh et al., 2006; Coe and Munoz, 2017; Neggers et al., 2005; Moore and Fallah, 2001). Similarly, the pre-frontal circuits involved in working memory are implicated in the eye movement control network responsible for fixation (i.e., periods between saccades where the line of sight remains relatively still) (Belopolsky and Theeuwes, 2009; Henderson and Choi, 2015; Henderson et al., 2015; Postle et al., 2006).

Accordingly, eye movements could be used a window into understanding the impact EFs have on information processing (Ayala and Heath, 2021; Bird et al., 2004; Diamond, 2013; Jobson et al., 2021; Pierrot-Deseilligny et al., 2003; Royall et al., 2002; Wang et al., 2016). Because EFs are related to various aspects of life (i.e., mental health, physical health, quality of life, school readiness, school and job success, violent/reckless behaviour, etc.) (Diamond, 2013), they serve as a crucial indicator for monitoring the progression of outcomes related to disease and rehabilitation in clinical populations (Ayala and Heath, 2021; Mazarandi and Gazerani, 2019; Wolf and Ueda, 2021), as well as for training and selection in educational and occupational settings (Brams et al., 2018; Peißl et al., 2018; Tien et al., 2015; Ziv, 2016). Notably, eye movement assays have been used to successfully probe EFs, however, the question remains how it transfers to ecological environments. To extend these findings, we need to specifically look at gaze behaviour (i.e., shifts in an individual's line of sight due to combined head and eye movements) rather than individual eye movements (i.e., the movement of eyes within the orbital fossa) to fully capture the way humans naturally interact with their environment. As such, investigating the relationship between gaze behaviour and task performance in natural and complex environments provides a framework for examining information processing and decision-making strategies that guide and influence goal-directed behaviour in real-life. Moreover, further experimentation that increases the task demands to match those experienced in more naturalistic settings is necessary to characterize the gaze behaviours and strategies people use to process and interact with the complexities of real-world tasks.

The establishment and validation of objective gaze metrics that are sensitive indicators of the perceptual and cognitive processes underlying task performance is the first step towards understanding how gaze behaviour supports complex goal-directed behaviour in naturalistic

settings. Therefore, the research in this thesis was conducted to systematically examine the utility of gaze behaviour as an objective indicator of information processing and skill performance during challenging tasks. This research provides a basis to identify how gaze behaviour mediates the efficient selection and processing of task-related information using complex tasks (i.e., tasks that involve sequences of sensorimotor decisions to be made via a series of sub-task goals) in a laboratory setting and a simulation environment. Specifically, the research was applied in the aviation context with the overarching aim to contribute to the development of a model that could inform pilot training and assessment.

The following section serves as an introduction to the current understanding of gaze behaviour and its applications to lab-based and natural environments. We start with a comprehensive review of eye movements and executive functions, followed by their application to clinical and occupational settings, and then present the overall goal of this thesis, along with specific objectives, research questions and hypotheses.

Section 2: Literature review

2.1 Oculomotor and gaze control

Broadly speaking, the two main functions of eye movements are to 1) stabilize the image on the fovea (i.e., fixational eye movements, vestibular-ocular reflex, and opto-kinetic nystagmus, and smooth pursuit), and 2) to direct the fovea to a target of interest in the periphery or at different depths (i.e., saccades and vergence) (Carpenter, 1988). The functional significance of stabilizing the object image on the fovea during fixation is directly tied to the anatomical features of the eye itself. Indeed, the receptors associated with early retinal processing in primates have a distinct pattern of organization (Rodieck, 1998). The cone photoreceptors responsible for photopic vision and high spatial resolution are densely situated at the center of the retina (i.e., the fovea). As you move away from the fovea (i.e., parafoveal vision, near peripheral vision, and peripheral vision) the density of cone receptors decreases exponentially and one's ability to perceive objects in detailed colour vision degrades (Valenko et al., 2019). In contrast, objects that fall closer to the periphery are characterized by low spatial resolution rod-mediated scotopic vision (Kandel et al., 2013). As such, humans have been reported to execute three to four saccades per second to ensure that areas of interest (AOIs) in the visual field benefit from high resolution foveal processing (Henderson, 2003; Yarbus, 1967). Individual fixations typically have an average duration ranging from 150-300 msec which supports encoding visual information gathered during each fixation 'snapshot' (Gegenfurtner, 2016). The visual information processed is dependent upon the six eye movements listed above that involve, 1) movement of the eyes within the orbit, and 2) the coordinated control of head movements that help shift and focus our line of sight (Kandel et al., 2013). Together, these movements comprise

gaze behaviour (i.e., coordinated eye and head movements) which is the primary method through which we naturally sample visual information in our environment (Lappi, 2016).

In natural environments, gaze behaviour, more than independent eye movements, provides input and guides our actions. For instance, version-vergence interactions are common and necessary to explore objects in our three-dimensional world. Additionally, the direction and stabilization of the fovea toward a particular AOI is achieved through the close coordination of eye movements and head movements. As such, the current thesis is primarily focused on examining gaze behaviour rather than any particular type of eye movement to fully capture the way humans naturally interact with their environment. Previous work was limited due to the nature of eye-tracking methodologies that required participants to complete paradigms where their head remained stationary via a chin/head rest or bite bar. However, with rapid technological advancements in eye-tracking technology we now have access to affordable, high-quality mobile eye and gaze-tracking systems that allow us to extend previous eye movement research to settings outside the lab.

2.1.1 The neural substrates involved in gaze behaviour

The act of moving the eyes and fixating an object of interest is an active process that involves the interaction of top-down (i.e., goal-directed) and bottom-up (i.e., saliency driven) processes (Land and Hayhoe, 2001; Land, 2006). These processes are primarily governed by a higher-level cognitive control mechanism that relies on executive functions (EFs) to prioritize the allocation of gaze – and by proxy, overt attention – to AOIs that provide crucial information, with respect to the task being performed (Land and Hayhoe, 2001; Land, 2006). Beyond the subcortical and brain stem circuitry involved in oculomotor control (i.e., superior colliculus,

pontine paramedian reticular formation, basal ganglia) several notable cerebral structures are involved in the generation and coordination of eye-head movements during task performance that are influenced by the competing bottom-up and top-down control mechanisms of overt attention. The areas that are most germane to the current dissertation include the lateral intraparietal area (LIP), supplementary eye field (SEF), frontal eye field (FEF), and dorsolateral prefrontal cortex (DLPFC) as they are commonly implicated in EFs as well as eye movement control (Collette et al., 2006; Cutsuridis and Yoshida, 2017; Freedman et al., 1996; Ghandi et al., 2008; Ghandi and Sparks, 2007; Knight and Fuchs, 2007; Martinez-Trujillo et al., 2003). Load-dependent increases in activation in the FEF, SEF, and DLPFC have been shown to be associated with increased demand placed on working memory, inhibitory control, and cognitive flexibility (i.e., top-down control) (Awh et al., 2006; Collette et al., 2006; Freedman et al., 1996; Ghandi et al., 2008; Ghandi and Sparks, 2007; Knight and Fuchs, 2007; Lee and Doris, 2014; Martinez-Trujillo et al., 2003). Other notable regions involved in bottom-up processing include the right hemisphere lateralized ventral frontoparietal network (i.e., the temporoparietal cortex and inferior frontal cortex) which are activated when behaviourally relevant stimuli appear that are salient or unexpected (Corbetta and Shulman, 2002). Moreover, seminal work has established a strong association between the neural basis of eye movement control and various neurological conditions which supports that eye movements could serve as a biomarker for diagnoses and prognoses of impairments in EFs (McDowell et al., 2002; Johnson et al., 2015a; 2015b; Pierrot-Deseilligny, et al., 2003; Heath et al., 2017; Wang et al., 2016). Though the use of individual eye movements provides a window into specific aspects of EFs, insight into the underlying neural correlates, and their association with task performance, there is much to explore in the generalization of these relationships to more elaborate eye movements and gaze patterns. As

such, research has slowly evolved to encompass the novel exploration of gaze behaviour (i.e., spatiotemporal gaze pattern) during tasks designed to capture the demands and complexities of vision in naturalistic settings.

2.2 Overview of executive functions

The performance of many day-to-day activities in our dynamic environment requires that we monitor and adjust our goal-directed actions to ensure safe and successful performance. Such actions are dependent on higher-level cognitive processes that are broadly referred to as executive functions, which are fundamental to our ability to engage in complex behaviours involving planning, problem solving, and decision-making (Diamond, 2013; Miyake et al., 2000). For instance, imagine driving a car and being stopped at a red traffic signal. If your intention is to move straight through the intersection, then you must withhold a response (i.e., accelerating through the intersection) when a green left-hand turn signal is presented. Indeed, executive functions allow you to selectively attend to and identify the nature of the stimulus, assert inhibitory control to avoid an inappropriate response (i.e., taking foot off the break), and maintain this information in working memory as you scan the rest of the environment for other task relevant information (i.e., pedestrians crossing the street). Clearly, each of these processes is crucial to the safe operation of a vehicle as an error/inefficiency in one, or more, may result in a car accident. More importantly, these processes represent four core dimensions that are proposed to underlie executive function, namely: 1) selective attention, 2) inhibitory control, 3) working memory, and 4) cognitive flexibility (Diamond, 2013).

Selective attention primarily functions as a selection mechanism that dictates which information will be processed with priority and which will be ignored (Oberauer, 2019).

Selective attention is governed by bottom-up (i.e., low-level, saliency driven) and top-down (i.e., high-level, goal-directed) processes that facilitate the selection and processing of relevant sensory stimuli (Diamond, 2013). Though the interaction of these competing processes is difficult to segregate and control entirely in naturalistic scenes, it is important to note that during goal-directed actions eye movements seem to be predominantly driven by top-down information hierarchies (i.e., scripts) rather than the ‘intrinsic salience’ of objects (Land & Hayhoe, 2001). The competition of these processes is evident more so in well-learned tasks than in novel tasks, where these top-down information hierarchies are first being established as learning progresses. As such these dominant top-down control processes actively draw upon maintained processing priorities and goal hierarchies from working memory and ignore task irrelevant stimuli via inhibitory control; which, collectively, guides the efficient sampling of information from our environment (Lavie et al., 2004; Posner and DiGirolamo, 1998). *Inhibitory control* is defined as the ability to deliberately inhibit one’s dominant, automatic, stimulus-driven or pre-potent response in favour of completing a more appropriate or goal-oriented response (Diamond, 2013; Miyake et al., 2000). *Working memory* is defined as the ability to update, maintain, and manipulate information in one’s mind that is no longer available in the sensory array (Diamond, 2013; Miyake et al., 2000). Lastly, *cognitive flexibility* is defined as changing perspectives or approaches to a problem by flexibly adjusting to new rules, demands, or priorities/goals (Diamond, 2013; Miyake et al., 2000).

Converging neuroimaging and lesion studies have shown that these executive function processes are robustly associated with prefrontal cortex activation. Specifically, increased activation in the DLPFC has been reported for tasks that involve goal-setting, inhibitory control, planning, sequencing, response-set formation, task shifting, verbal and spatial working memory,

self-monitoring, and self-awareness. The orbitofrontal cortex activation has been associated with initiation of social and internally driven behaviours, as well as the inhibition of inappropriate behavioural responses. The medial prefrontal cortex was found to mediate tasks that require attention, inhibitory control, planning and memory, and the anterior cingulate cortex is involved in behavioural monitoring and error correction (Bird et al., 2004; Diamond, 2013; Jobson et al., 2021; Mansouri et al., 2019; Royall et al., 2002; Truelle et al., 1995). Though there has been a continued debate whether these EF processes rely on the same or similar underlying mechanism, most researchers agree that they are somewhat distinct and contribute differentially to performance of complex tasks (Diamond, 2013; Miyake et al., 2000). As such, there is a need to build on previous work that examined the core domains of executive function in isolation in highly controlled laboratory settings by increasing task complexity to parallel the demands seen in real world scenarios. Therefore, approaches and tests challenging EF paired with gaze behaviour could be useful in revealing the efficient and/or inefficient information processing mechanisms associated with task performance in naturalistic settings. Such knowledge would be useful in assessing patients with neurological deficits to predict their ability to function in daily life, or alternatively, in assessing skill performance development and proficiency in complex occupations (i.e., aviation, medicine, military operations).

2.3 Eye movements and their application to the real world

Research has shown an intimate link between gaze shifting, cognitive, and motor processes (de Brouwer et al., 2021). For instance, gaze behaviour has been shown to reflect the allocation of attention in the environment (Moehler and Fiehler, 2018) and ability to resist distracting stimuli (Jonikaitis et al., 2019). This is particularly evident in the gaze behaviours

examined during the performance of a neurocognitive assessment task in healthy adults (i.e., Tower of London) (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012; Rowe et al., 2001). Several studies using this assessment tool have consistently demonstrated that the amount of time that gaze was directed toward particular AOIs was biased toward task-critical regions that required more information processing resources to elaborate and maintain solution sequences for successful task performance. Additionally, some of the latter work demonstrated how gaze was used to differentially support increases in working memory load and problem-solving processes (Parker et al., 2020), as well as identified how putative changes in gaze behaviour were indicative of erroneous information processing processes associated with poor performance (Hodgson et al., 2000; Hodgson et al., 2002; Huddy et al., 2007).

The relationship between gaze, cognitive and motor processes has also been particularly evident in limb movement research that has demonstrated how vision is specifically used to coordinate motor responses, especially when gaze needs to be directed to encode information needed for the successful execution of the task. This relationship is notable in several tasks involving pointing (Abrams et al., 1990; van Donkelaar et al., 2004), object manipulation (Hayhoe and Ballard, 2005; Hayhoe et al., 2003; Land and Hayhoe, 2001), walking (Hollands et al., 2002), and other day-to-day activities (i.e., making a sandwich or cup of tea, or driving) (Land & Hayhoe, 2001; Leonhardt and Wanielik, 2017; Martin et al., 2017), which have reliably demonstrated that eye movements are typically initiated towards the object of interest 40-60 ms prior to movement onset/object interaction (Biguer et al., 1982; Land et al., 1999). These fixations are linked to where participants are manipulating/interacting with objects or linked to forthcoming obstacles and landing sites where objects are subsequently grasped, moved around, intercepted, and/or placed, respectively (Johansson et al., 2001; De Nocera et al., 2007;

Panchuck and Vickers, 2006; Vickers, 2007). Taken together, much research has conclusively shown that the visual information gathered through sequences of fixations has a direct influence on the planning and execution of goal-directed actions.

One of the more common accounts of the characteristic gaze behaviours that emerge during goal-directed actions is that eye movements seem to be driven much more by the top-down information hierarchies (i.e., scripts) rather than the ‘intrinsic salience’ of objects (Land & Hayhoe, 2001). For instance, the the perception of pictorial depth cues (i.e., top-down) in an image may influence the perception of the size of an object (i.e., bottom up) in a manner that rescales the stimuli as a function of how far the appear to be. In this example, farther objects would appear to be larger, while closer objects would appear to be smaller (Gregory, 1968). Therefore, current literature supports that eye movements provide a useful probe for understanding the top-down information processing (i.e., EF-driven) mechanisms that support successful task performance. Specifically, analysis of gaze behaviour provides insight into processes that help to explain how individuals select task-pertinent information by fixating at the relevant locations at the right time to effectively plan and execute highly skilled movements. Not surprisingly, the analysis of eye movements and gaze behaviour has become a prominent area of interest in aviation and medical related research, which aims to examine the complex relationship between eye movements and motor strategies that go beyond lab-based paradigms (Martinez-Marquez et al., 2021).

Several metrics derived from eye movement analyses have been identified and developed to help probe different cognitive, emotional, and physiological states, which can be used to gain a more in-depth understanding of the information processing mechanisms that underlie skill

performance (Land, 2006; Land & Hayhoe, 2001; Hodgson, 2019; Rowe et al., 2001).

Traditional approaches to analyzing gaze behaviour involve discrete measures related to the total proportion of fixation time (i.e., dwell time), frequency or average durations directed to different regions in the environment, which provide insight on where and how attention was allocated (Galholt, 2014; Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012; Rowe et al., 2001). Specifically, areas for which the displayed information changes frequently or that is of critical importance to the task are likely to receive a larger proportion of the total dwell time as well as a higher frequency of fixations (Brams et al., 2018; De Nocera et al., 2007; Peißl et al., 2018; van de Merwe et al., 2012; Ziv, 2016). The average duration of the fixations can also serve as an indication of information processing efficiency. For example, increasing fixation durations have been linked to an increase in the cognitive load associated with the task, and have also been shown to be characteristic of the longer time required to process the information being fixated in novice or inexperienced pilots (Brams et al., 2018; Galholt, 2014; Peißl et al., 2018; Ziv, 2016). Moreover, scan path length can serve as an indication of high workload conditions, where attentional resources become strained, and thus, saccade amplitudes and total scan path lengths become shorter to focus on a smaller region of the task environment or fewer AOIs (Galholt, 2014). Notably, these are more discrete measures of gaze behaviour that are often time-averaged operations and fail to make use of the information regarding the patterns and sequences of the visual scan sequence. Examining gaze pattern sequences may provide additional insight regarding the informational processes that support the progression of goal-directed actions.

In accounting for the limitations of traditional gaze measures, more computationally heavy measures have been developed and applied to identify global patterns of behaviour that emerge over the course of task performance (Brams, 2018; Galholt, 2014; Peißl et al., 2018;

Shiferaw et al., 2019; Ziv, 2016). Gaze entropy represents an information theory measure that captures the complexity of the global patterns within gaze behaviour (Shannon, 1984; Shiferaw et al., 2019). Namely, stationary gaze entropy (SGE) and gaze transition entropy (GTE) have emerged as objective indices of gaze dispersion and fixation sequence complexity, respectively (Shiferaw et al., 2019). It is important to note that the values provided by entropy measures are relative to the number of AOIs employed and are task specific. To aid in generalizing findings across experiments despite differences in the number of AOIs used, it is recommended that any reported entropy values are normalized between 0 and 1. In any case, a high entropy value (0.7-1) tends to be more reflective of an exploratory mode of gaze allocation (i.e., less attentional bias, fixating several areas that are relevant and/or irrelevant to the task at hand), while a low entropy (0-0.03) is indicative of a focal mode of attention (i.e., structured, and biased attention toward fewer AOIs). Understanding the context of the task goals being performed is necessary for correct interpretation of these values. For example, higher values may be more advantageous for monitoring tasks typically associated with driving and aviating in dynamic environments, whereas lower values may be more advantageous for managing routine or emergency situations where the allocation of gaze is most effective in supporting task performance when directed toward a select few number of AOIs. Although further research is required to fully understand how changes in cognitive load, level of expertise, type of task, and other factors influence gaze entropy measures, they have been shown to be a useful metric that provides additional information about the global spatiotemporal properties of gaze behaviours associated with information processing and goal-direction actions (Brams et al., 2018; Shiferaw et al., 2019).

Section 3: Research aims, questions and hypotheses

3.1 Research Aims

The overarching goal of this thesis is to advance the fundamental understanding regarding the utility of gaze behaviour as an indicator of information processing during the performance of complex tasks encountered in occupational settings. Based on the framework outlined below (Figure 3.1), this research seeks to provide a basis for examining whether gaze behaviour could shed important insight on the efficiency and optimization of information processing that supports complex task performance. Accordingly, studies were designed to determine if changes in gaze behaviour are associated with information processing demands, specifically, task demands (i.e., sensorimotor difficulty and cognitive load) and the individual's resource capacity (i.e., flight hours/expertise). Importantly, this work also examined the generalizability of the gaze measures across various environmental contexts.

The general approach to guide the experimental work described in this thesis is captured by the information processing and skill performance framework depicted in Figure 3.1. Here, *information processing* is defined as an individual's ability to select and process sensory information in their environment to generate a required goal-directed behaviour. It is important to note that this involves an individual's ability to collectively carry out the necessary sensory, perceptual, cognitive, and motor processes associated with a given task. Information processing load is influenced by the interaction between *task requirements*- the physical and temporal structure of the task (i.e., configuration, duration, and type; aspects that impact sensorimotor difficulty and cognitive load), *individual resources*- which can be influenced by personal experiences (i.e., education, hours of practice, neurological deficits, genetics), and *environmental context*- where/how the task is being performed (i.e., simulator vs. real plane, presence of

instructor, presence of co-pilot). In a very basic sense, the framework can be understood in a similar manner to the way a flashlight is used in a dark room (Posner, 1980). Gaze behaviour acts as the light, flashing on objects in the dark to acquire detailed information that may be relevant to the task at hand. Though the environment, individual, and task provide crucial context to what makes specific areas of interest more important than others, the information processing mechanisms that make up the handle of the flashlight are critical to guiding where the flashlight is directed next and for how long to adequately acquire and update all necessary information relevant to completing the task. EFs are the key information processing component of interest as its role in balancing the top-down and bottom-up processes guiding information acquisition play a particularly prominent role in how well an individual performs a task. I propose that gaze behaviour serves as a non-invasive probe that provides insight into information processing and can help understand skill performance. Because task performance can vary on a spectrum from poor (i.e., failure or completing a task in an inefficient manner) to exceptional (i.e., completing a task effectively and efficiently), gaze behaviour may help discern the differences in information processing mechanisms that distinguish individuals that excel from those who do not under similar task conditions.

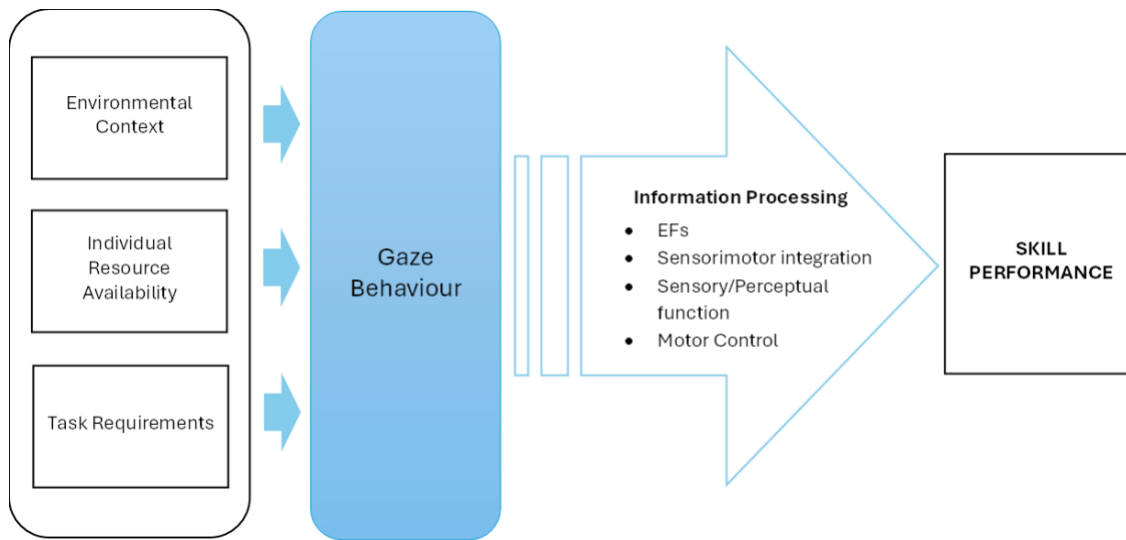


Figure 3.1- An overview of information processing and the contributing factors underlying skill performance. Gaze behaviour refers to eye movements which serve to acquire visual information. Since the area of high resolution (i.e., spatial acuity) is limited, examining gaze behaviour highlights where and when attention is being deployed based on environmental factors (i.e., simulated/real environments, presence of evaluators), individual resource availability (i.e., level of experience), and task requirements (i.e., sensorimotor demands and cognitive load). The manner through which gaze is selectively deployed is driven by competing bottom-up and top-down processes that are governed by Executive Functions (EFs). Understanding the appropriate balance of top-down and bottom-up processes to meet the needs of the environment, individual, and task has critical implications for skill performance and learning.

3.2 Overview of studies: Research questions and hypotheses

Study 1

The primary question of this thesis – how effective is the assessment of gaze behaviour as a proxy for information processing during complex task performance? – was first addressed in study 1, which was particularly imperative for developing and establishing a comprehensive set of ecologically valid gaze behaviour metrics. This first study was designed to examine a comprehensive set of gaze metrics to determine the spatiotemporal aspects of gaze that were most sensitive/effective in characterizing changes in information processing due to cognitive load manipulations. The purpose of starting with a well-established and highly controlled neurocognitive assessment task (i.e., Tower of London: TOL) was to derive a comprehensive set of gaze metrics using a paradigm that has previously demonstrated alterations in information processing via gaze metrics, neuroimaging, and behavioural performance; albeit, with more simplistic problem sets (Hodgson et al., 2000; Kaller et al., 2009; Nitchke et al., 2012; Rowe et al., 2001). The specific research questions and hypotheses were:

1. Does gaze behaviour reflect increased cognitive load due to increased visuospatial working memory demands?

Hypothesis: Gaze behaviour will reveal attentional biases which will be modulated by increases in cognitive load. Specifically, results will reveal that information from task-relevant AOIs is prioritized, especially during higher cognitive load conditions

2. Which gaze outcome measures reveal deficits in information processing during task performance?

Hypothesis: Dwell time changes that reflect a lack of attentional biases prioritizing task-relevant information will be associated with less efficient task performance.

Study 2

Study one provided support for the use of gaze behaviour as a probe into information processing during challenging neurocognitive tasks. However, the utility in examining such a comprehensive set of gaze metrics in real-life tasks has not been previously investigated, particularly in low-time pilots. Therefore, the second study tested participants who were familiar with using a Microsoft Flight Simulator game while they performed a series of landing tasks that varied in sensorimotor control demands (i.e., presence of significant cross wind) using a PC-based flight simulation program. Additionally, a novel cognitive tunneling analysis was developed to determine its sensitivity in identifying changes in poor cockpit scanning behaviour in low-time pilots as sensorimotor demand increased.

The specific research question and hypothesis was:

1. What is the effect of increasing task difficulty (i.e., turbulent weather conditions) on gaze behaviour and landing performance?

Hypothesis: Increasing task difficulty will be associated with an increased gaze bias toward task relevant AOIs that provide critical information for task performance, which will be associated with level of success.

Study 3

Studies one and two indicated that alterations to task requirements are associated with changes in information processing and gaze behaviour across a neurocognitive task and flight simulation

landing task. However, both tasks were still conducted in a highly controlled laboratory setting using a desktop PC and the EyeLink II eye tracker, which constrained the movements and interactions that would normally occur during task performance in ecological environments. In this study, the validation of gaze behaviour findings was examined in a fully immersive flight training device (AL 250, ALSIM) using the AdHawk Mindlink portable eye tracker while sensorimotor demand was manipulated (i.e., presence of strong cross wind). This work further expanded on study 2 by breaking down the landing scenario into three different stages that are known to be associated with different task goals.

The specific research questions and hypotheses were:

1. Are the effects of task difficulty on gaze behaviour and landing performance observed in a desktop-based flight simulation device generalizable to a high-fidelity flight simulation device?

Hypothesis: In line with previous work, the gaze patterns associated with task performance during a flight landing task will generalize from a desktop-based flight simulation device to a high-fidelity flight simulation device. It is expected that an increase in task difficulty will be associated with an increased bias toward task relevant AOIs that provide critical information regarding task performance.

2. Is gaze behaviour associated with pilots' subjective report of situation awareness?

Hypothesis: The more attentive pilots are to task relevant AOIs as reflected by their gaze behaviour, the higher their subjective situation awareness will be.

Study 4

The first three studies provided evidence showing the utility of gaze behaviour as a proxy for information processing across changes in task requirements in both low-fidelity (PC desktop) and high-fidelity (ALSIM simulator) environments. However, two key methodological factors requiring further investigation became apparent. First, the processing and analysis of AOI-based gaze behaviour requires tremendous effort in manually defining each AOI for every trial and every participant. Fiducial markers serve as a method to address this limitation, but their appearance is distracting in natural environments and may interfere with natural gaze behaviours and task performance. Therefore, this study examined the extent to which original visible markers and novel invisible markers impacted participant task performance and gaze behaviour during simulated flight. Second, gaze behaviour during task performance is known to be impacted by expertise. Novices tend to exhibit scan patterns that are less expansive and show less of a bias between task relevant and irrelevant AOIs (Brams et al., 2018). Thus, two groups of pilots were included to further assess the extent to which pilot experience influenced the degree to which task performance and gaze behaviour were altered by fiducial marker type; thereby examining the impact of individual resource availability- assumed by differences in expertise- on information processing. The specific research questions and hypotheses were:

1. Does flight experience influence the extent to which the original visible fiducial markers are more distracting than the novel invisible fiducial markers?

Hypothesis: Visible fiducial markers are expected to negatively impact flight performance, increase subjective distraction scores, and demonstrate gaze behaviour patterns that are indicative of an increase in exploratory behaviour and increased bias toward task-irrelevant AOIs (i.e., markers); a finding that evinces the original markers to

be more distracting than the proposed novel markers. This effect is hypothesized to be more prominent in novice pilots than licensed low-time pilots.

2. What impact does flight experience (i.e., novice versus licensed low-time pilot) have on gaze behaviour and task performance?

Hypothesis: In line with previous work, the gaze patterns associated with expert pilot behaviour are suggested to show indications of capturing more information from several task-relevant AOIs while being more effective at ignoring task-irrelevant AOIs. Novices on the other hand may not attend to as many task-relevant AOIs during task performance and may also have an equal tendency to attend to task-irrelevant AOIs.

Study 5

Studies one and two provided evidence that gaze behaviour is a robust measure for assessing changes in information processing as sensorimotor demands increase during simulated flight. Notably, studies three and four suggest that licenced low-time pilots are particularly effective at using gaze to selectively attend to several task-relevant AOIs and are able to do so with minimal impact on flight performance when faced with increasing sensorimotor demands and distracting stimuli (i.e., visible fiducial markers). Accordingly, to provide a means to further challenge for this group of recruited pilots, I designed a dual-task paradigm to evaluate the utility of gaze behaviour in examining changes in information processing due to increased cognitive load across two phases of flight (i.e., cruise, landing). I chose to introduce an auditory oddball paradigm as the secondary task to monitor auditory inputs as a means to mimic the modality of Air Traffic Control (ATC) calls that pilots often must attend and respond to during flight.

The specific research question and hypothesis was:

1. What is the effect of increasing cognitive load on gaze behaviour and flight performance?

Hypothesis: Gaze behaviour will show an increased bias toward task relevant AOIs in the high cognitive load condition compared to the low cognitive load condition. Gaze entropy measures may provide additional insight into how well individuals will perform on the secondary task.

Section 4 – Gaze Behaviour: A window into distinct cognitive processes revealed by the Tower of London test

A version of this chapter has been published in Vision Research (Ayala et al., 2022) and the author retains the right to include this article in their thesis or dissertation.

Ayala N, Zafar A & Niechwiej-Szwedo E. Gaze behaviour: A window into distinct cognitive processes revealed by the Tower of London test. *Vision Research*. 2022 Oct;199:108072. doi: 10.1016/j.visres.2022.108072. Epub 2022 May 24. PMID: 35623185.

4.1 INTRODUCTION

Gaze behaviour refers to the pattern of fixations and eye movements which enables sampling of high-resolution visual information that is crucial for perceiving and interacting with our environment. Seminal work has demonstrated that gaze behaviour is highly task-dependent and tightly linked to the underlying perceptual, cognitive and motor processes (Castelhano, Mack and Henderson, 2009; de Brouwer, Glanagan, and Spering, 2021; Flanagan, Terao, and Johansson, 2008; Land and Hayhoe, 2001; Peltz, Hayhoe, Loeber, 2001; Yarbus, 1967). As such, the analysis of gaze behaviour during complex tasks provides a promising and non-invasive method to examine how specific eye movement patterns relate to various aspects of cognition and action. Notably, most research to date has examined the association between aspects of gaze behaviour and subsequent goal-directed actions using relatively simple problem-solving paradigms (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012; Rowe et al., 2001) that do not match the complexity of those used in clinical or real-world applications (Shallice, 1982). Furthermore, there is a lack of reliable objective measures that effectively characterize

task performance during these challenging tasks. Therefore, the present investigation sought to examine the utility of using gaze-based metrics to characterize the performance proficiency of healthy adults during a challenging problem-solving task, and to determine how changes in gaze behaviour support the increasing demands placed on planning and working memory.

Gaze behaviour has been analyzed in various lab-based paradigms that specifically rely on executive functions (i.e., working memory, response inhibition, and attentional set-shifting) (Hodgson et al., 2000; Kaller et al., 2009; Martin et al., 2017; Veneri et al., 2012). These higher-level cognitive processes provide the foundation for goal-directed behaviour (Diamond, 2013). Notably, optimal performance of most goal-directed behaviours tends to rely on an individual's capacity to plan and problem solve, which entails the ability to think ahead, set goals, and anticipate the consequences of actions prior to their execution (Goel and Grafman, 1995). One of the most common paradigms used to examine complex visuospatial problem-solving abilities is a disk-transfer task called the Tower of London (TOL), which was originally created to assess planning deficits in patients with frontal lobe lesions (Shallice, 1982). The goal of this task is to rearrange an array of balls or discs placed on pegs (i.e., workspace) such that the pattern matches the arrangement of a second array of balls (i.e., goal space) using the minimum number of moves specified in the instructions. Task difficulty is dictated by several factors including the minimum number of moves required to complete the goal, goal hierarchy (i.e., ambiguity of the sequence of final moves derived from the configuration of the goal state), and search depth (i.e., number of intermediate 'shunting' moves) (for review see Kaller, Unterrainer and Stahl, 2012). Participants are instructed to carefully plan the full sequence of moves prior to executing the task (i.e., using a computer mouse to execute the planned sequence of moves); thus, dissociating the cognitive planning and solution elaboration components of the task from action execution and verification

processes (Kaller et al., 2009; Rowe et al., 2001). Indeed, failure to plan the solution ahead of time will likely result in initial moves blocking subsequent moves, and the execution will require more moves than the minimum (i.e., optimal) number specified.

Several studies have addressed the spatial distribution and temporal sequence of natural eye movements during performance of the TOL task (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012). For instance, Hodgson et al. (2000) reported that planning is organized in discrete phases, with an initial assessment of the goal space which provided the necessary information to create an internal representation of the goal state that subsequently led to solution elaboration. Solution elaboration was then completed while participants predominantly focused their attention on the workspace, a result made evident by the disproportionately longer dwell times in this area of interest compared to the goal space. Moreover, data demonstrated that efficient planners biased their gaze toward problem-critical balls in the workspace (i.e., balls that must be moved prior to others in order to solve the problem) whereas participants who made errors did not selectively direct their gaze towards these problem-critical locations. Building on this, more recent research has demonstrated that the use of eye movement data allows for reliable identification of different stages of problem-solving processes that occur in the initial planning and execution stages of the task, as well as differential task demands on a trial-by-trial basis (Kaller et al., 2009; Nitschke et al., 2012). For example, the duration of the first two fixations during the planning interval is relatively stable regardless of experimental manipulation, which was suggested to reflect the internalization of the goal space (Kaller et al., 2009). Additionally, the last fixation duration was reported to reflect solution elaboration as it was exclusively influenced by the experimental manipulations (i.e., minimum number of moves, search depth and tower configuration) (Kaller et al., 2009; Nitschke et al., 2012). Collectively, these studies

demonstrated that when the requirements of a task are as highly specific as they are in the TOL, examining eye movements during planning and execution can reveal discrete cognitive processes and strategies used to facilitate task performance.

The traditional gaze metrics explored in recent work provide adequate detail for the identification of eye movement patterns that support neurocognitive performance in simple TOL problems (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012). However, the planning and decision-making strategies used to complete more demanding TOL problems may require more advanced methods of gaze behaviour analysis. Gaze entropy is an objective and novel approach to gaze analysis that may provide additional information about the spatiotemporal properties of gaze behaviours associated with cognitive processing and goal-directed action (Shiferaw et al., 2018). Notably, entropy (measured in bits of information) has been defined as the average information or uncertainty associated with choice (Shannon, 1984), which can be used to quantify the complexity of a system (Batty et al., 2014). In the context of gaze behaviour, where and when we fixate relative to a set of specified areas of interest (AOI) (i.e., choices) in the environment governs the complexity of an individual's gaze pattern. Accordingly, two complementary measures of entropy have emerged as objective indices of gaze complexity, namely stationary gaze entropy (SGE) and gaze transition entropy (GTE) (for review see Shiferaw et al., 2018). SGE is a measure of gaze dispersion during a defined time interval. It is calculated by applying Shannon's entropy equation (Equation 1, see Methods) to the probability distribution of a sequence of fixations across the AOIs. In this case, the complexity of a gaze pattern reflects the number of different AOIs fixated, and how the pattern of fixations is distributed across the AOIs (Shiferaw et al., 2018). Thus, a higher entropy value indicates larger gaze dispersion and represents more unpredictability or randomness in the

distribution of fixations. On the other hand, GTE builds on SGE to further examine the visual scan pattern complexity/unpredictability by considering the rate of transitions between AOIs (Shiferaw et al., 2018). GTE assumes that subsequent fixation locations in a scan pattern are better predicted from current locations through conditional probability rather than overall probability (Wiss et al., 1989). In other words, the probability of the next fixation location being in a given AOI is generated from information of the current AOI and the previous AOI. GTE is calculated by applying the conditional entropy equation (Equation 2, see Methods) using Markov chain matrices of fixation transitions. In this instance, higher entropy suggests a less structured or a more random pattern of sequential scanning behaviour. Together, SGE and GTE could provide a quantitative method of characterizing gaze behaviour related to cognitive processes and decision-making strategies in complex environments.

The present work examined gaze behaviour and gaze entropy in healthy adults while they performed a standardized version of the TOL task. Notably, the present investigation differs from previous work in three important respects. First, the current work employs the Freiburg Version of the Tower of London (TOL-F) paradigm (Kaller et al., 2012). This version of the TOL task is used in clinical neuropsychology and enables the assessment of gaze behaviour during more demanding higher-level problems compared to the ones used in previous studies (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012). The most challenging problem set employed here requires a minimum of 6 moves to accomplish the task goal, whereas previous studies used problem sets with 3 moves or fewer (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012). It is important to extend previous findings to these more difficult problem sets in order to establish the spatial distribution and temporal sequence of eye movements that underlie optimal planning and problem-solving performance in tasks that are used in current

clinical/neuropsychological assessment tools. Second, using more difficult problem sets allowed us to conduct an in-depth analysis of non-optimal trials to determine whether putative changes in gaze behaviour could distinguish optimal and non-optimal performance. Third, the current work introduces and examines the utility of entropy measures as an objective method to analyze cognitive load and neurocognitive performance in a standardized neurocognitive assay. It is important to note that previous work using entropy metrics in naturalistic settings has suggested that SGE is more susceptible to the influence of visuospatial demand (i.e., stimulus properties), whereas GTE is more reflective of inherent task difficulty and increased cognitive load (Shiferaw et al., 2018). However, more research is required to specifically examine how changes in cognitive load influence both SGE and GTE, which can be directly addressed through the TOL task.

To our knowledge, the current study is the first to systematically examine the application of entropy measures in conjunction with traditional gaze analyses during complex (i.e., TOL problem sets involving 3 or more moves) visuospatial problem solving in healthy adults. In line with previous studies employing traditional gaze (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012) and entropy metrics (Shiferaw et al., 2018), we hypothesized that an increase in task difficulty (i.e., increased minimum number of moves to complete the task) will result in a workspace-specific increase in dwell time and an increase in GTE during the planning stage of the task. The expected dwell time and GTE findings are proposed to underlie a greater need for more complex scanning, solution elaboration and mental manipulation of the visuospatial arrays in visuospatial working memory to successfully execute longer sequences of solutions. Additionally, we hypothesize that SGE will be equivocal across the task difficulty conditions during the planning stage of the task. Alternatively, if SGE is modulated by task

difficulty during the planning stage, then it can be concluded that SGE is another indicator of cognitive load and not merely a measure of perceptual demand. Gaze behaviour has not been examined in detail during the execution stage of the task; therefore, we aim to characterize the spatiotemporal gaze patterns (with both traditional and entropy analyses) associated with the execution of challenging visuospatial problem-solving. Furthermore, it has been suggested that the bias in fixations toward the workspace during planning is crucial for optimal task performance, and that efficient planners focus their gaze on problem critical balls (Hodgson et al., 2000). Notably, the absence of such a gaze bias toward the workspace has been shown to represent a failure in problem solving processes (i.e., inability to maintain task-relevant information in working memory) (Hodgson et al., 2002; Huddy et al., 2008). As such, we hypothesized that non-optimal trials would not be associated with a workspace-specific increase in dwell time during the planning stage of the task but would be linked to an increase in SGE (i.e., more dispersed fixation distribution) and GTE (i.e., unstructured/unpredictable gaze pattern). Together, these findings would suggest that non-optimal trials are a product of faulty encoding and/or maintenance of an internal representation of the goal state during the planning stage of problem solving.

4.2 METHODS

4.2.1 Participants

Twenty-seven participants were recruited from the student population at the University of Waterloo (8 males: age range 18-31 years, mean=22 years old, SD= 3 years). All participants had normal or corrected-to-normal vision. All participants also indicated that they had not been previously diagnosed with a neuropsychiatric/neurological disorder or learning disability.

Participation in the study was voluntary, and participants did not receive compensation for their involvement. The study's protocol was approved by the University of Waterloo Research Ethics Board Committee, and consent was obtained prior to beginning the experiment.

4.2.2 Materials and procedures

Participants first completed a visual screen including a visual acuity test using the Bailey-Lovie chart and a stereoacuity test using the Randot Stereo test (Stereo Optical Company, Inc). Participants then sat in a height-adjustable chair with their chin placed on a chinrest for the duration of data collection. Visual stimuli were presented on a 19-in. CRT monitor (85 Hz refresh rate, 1024x768 pixels, View Sonic) at a viewing distance of 80 cm, with the center of the monitor aligned with the participant's midline. Eye position was recorded using the EyeLink II eye-tracker (SR Research, Ontario, Canada) sampling at 250 Hz. Prior to data collection, a nine-point calibration was performed. An immediate follow-up validation of calibration accuracy was conducted to verify that the error was $<1^\circ$ for each point in the calibration grid. Stimulus presentation and behavioural data acquisition were controlled using the Vienna Test System (Schuhfried, Moedling, Austria) running on a laptop (ThinkPad T430 Lenovo, Intel Core i5-3230M Processor– 3.20GHz, 3MB Cache, 1600 MHz).

Participants were tested in a single session (approx. 15-20 minutes). The Freiburg Version of the Tower of London (TOL-F) task consisted of a standardized set of 28 problems (i.e., trials) showing two arrays (6.42° by 5.35°) of three coloured balls positioned on three columns of varying heights (**Figure 4.1**). The two arrays of balls were located in the upper (i.e., goal space) and lower (i.e., workspace) visual fields. The leftmost location in each array had the tallest column with enough space to fit a maximum of three balls, the middle column could fit

two balls, and the right column had space for only one ball. The minimum (i.e., optimal) number of moves required to complete each problem was shown to the left of the workspace. The number of moves increased incrementally across the trials from 3 to 6 moves and contributed to the difficulty level of the task (i.e., easiest=3 moves; hardest=6 moves). The 28 problems consisted of 4 trials at level 3 (i.e., 2 trials with an intermediate ‘shunting’ step and 2 trials without an intermediate ‘shunting’ step), and 8 trials at levels 4, 5, and 6 (i.e., all trials included at these levels involved an intermediate ‘shunting’ step). ‘Shunting’ is a necessary subgoal location that a ball is temporarily moved to when solving a TOL problem (Hodgson et al., 2000).

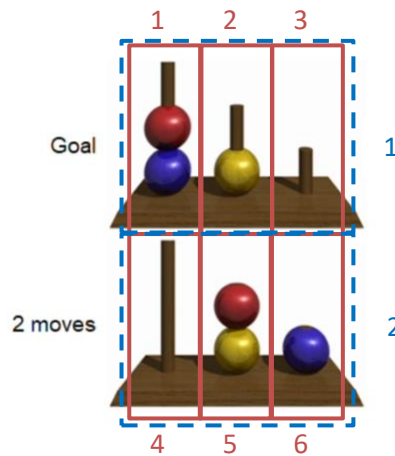


Figure 4.1. Illustration of the visual stimuli employed in the Freiburg Version of the Tower of London task. Goal space is displayed in the top array and workspace is displayed in the bottom array. The minimum number of moves required for the problem is indicated to the left of the workspace. The blue boxes represent the two main areas of interest used in the traditional gaze analysis. The orange boxes represent the six main areas of interest used in the entropy analysis.

Prior to starting the experiment, participants were introduced to the TOL-F task and provided with instructions concerning how the balls could be moved from one column to another, and that the balls could not be placed directly underneath another ball without moving the obstructing balls to an alternative location. It was also emphasized that participants should mentally plan out the entire sequence of moves required to solve each problem prior to executing the solution (i.e., moving the balls via a computer mouse). All participants were given a block of practice trials using the same three-ball arrangement used in the experiment with simple planning task loads (i.e., a set of 3 simple two-move problems). Eye movements were not recorded during these practice sets.

Visual stimuli and trial sequence for the experimental protocol were the same for all participants. At the start of each trial, participants were asked to look at a red dot marked on the monitor in order to standardize initial eye position. The trial was then manually initiated after a drift correction was completed by the examiner. The problem set was displayed until the participant solved the task or for a maximum of 60 seconds - whichever came first. Trials were categorized as optimal (i.e., participant solved the problem in the prescribed number of moves), non-optimal (i.e., participants used more moves than 'optimal') or time expired error (i.e., participants failed to solve the problem in 60 seconds).

4.2.3 Data reduction

Eye movement data were analysed offline using the eye tracker's Data Viewer software (ver 1.8; SR Research, Ontario, Canada). Eye-movement traces were visualized by the experimenter and played back at slowed speed superimposed over the image displaying the problem set. Eye movements were detected using a saccade detection algorithm implemented in

Data Viewer: $30^\circ/\text{s}$ velocity threshold and $8000^\circ/\text{s}^2$ acceleration threshold. Data Viewer identified fixations as pauses between saccades. The current study focused on primary saccades, thus microsaccades ($<1^\circ$) were excluded from analysis (Martinez-Conde et al., 2013). To ensure microsaccades were not an influential factor, the post-hoc analyses listed below for the traditional gaze measures (see *Statistical analysis*) were also exclusively completed for microsaccades which did not reveal any significant effects or interactions. Trials with missing data (i.e., loss of signal $>30\%$) ($<9\%$ of trials) and outliers (i.e., >1.5 the interquartile range above the third quartile) ($<10\%$ of trials) were removed. Due to insufficient trials, time expired error trials were also excluded from analysis, which accounted for less than 2% of trials.

4.2.4 Performance analysis

Performance-dependent variables included accuracy (%), initial planning time (sec) and movement execution time (sec), which were derived from the Vienna Test System (Schuhfried, Moedling, Austria). Accuracy was defined as the percentage of trials out of the total number of trials where the problem was correctly solved in the minimum number of moves (i.e., optimal trials). Initial planning time was defined as the duration from problem set display onset to the first mouse click on a ball, which constitutes the planning stage. Movement execution time was defined as the duration between the first mouse click and the completion of the problem (i.e., last mouse click), which represents the execution stage. Initial planning time and movement execution time were subsequently used to segregate the analysis of gaze data (traditional and entropy) into their respective planning and execution stages.

4.2.5 Traditional gaze analysis

The screen was divided into two areas of interest (AOI): goal space and workspace (**Figure 4.1**). Fixations on the screen above and below the fixation cross were assigned to the AOI of goal space and workspace, respectively. Gaze-dependent variables for each AOI included: dwell time (%), fixation rate (fixations per second), fixation duration (ms), and total scan path length (°). Dwell time was defined as the total duration spent within a given AOI during the planning or execution intervals, which was converted to a percentage (i.e., with respect to total trial time). Fixation rate was defined as the number of fixations that occurred relative to the total time spent during the planning or execution intervals, in the workspace and goal space respectively. Fixation duration was defined as the average duration of all fixations within a given AOI that occurred during the planning or execution intervals. Last, scan path length was defined as the sum of saccade amplitudes within each respective AOI that were completed during the planning or execution intervals.

4.2.6 Entropy analysis

The task environment was discretized within Data Viewer by organizing the goal- and workspace into six areas of interest (AOIs) (**Figure 4.1**). The AOIs were defined using a 3x2 grid with 3 regions in the goal space and 3 in the workspace where each AOI was centered on the column where balls were manipulated. Fixations found outside this 3x2 grid were defined as a non-area of interest and excluded from the analysis. Eye fixations in the six AOIs were assigned a number from 1 to 6 indicating the AOI where the eyes fixated. A sequence of fixation locations was then generated for the planning and execution phases of each trial. Custom scripts were

written in Python to compute both stationary gaze entropy (SGE) and gaze transition entropy (GTE).

SGE represents the overall uncertainty in the AOI that eyes may be fixated on at any given moment; a higher SGE value indicates that fixations are more spatially “dispersed” across the AOIs during task completion. In order to compute SGE given a sequence of fixation locations, first a vector, $\mathbf{V}$, of length 6 was produced, where $\mathbf{V}_i$ was the total number of fixations at AOI i . $\mathbf{V}$ was then divided by the total number of fixations in the sequence, so that $\mathbf{V}_i$ was the probability of a fixation landing at AOI i . Finally, SGE was computed using Equation 1 (Shannon, 1948), using the probability vector $\mathbf{V}$.

$$H_{SGE}(V) = - \sum_{v \in V} v \cdot \log(v)$$

Equation 1

GTE represents the overall uncertainty in knowing the next fixation location, given the current fixation location; a higher GTE value indicates that scan-paths were more unpredictable (i.e., less structured) during task completion. In order to compute GTE given a sequence of fixation locations, a 6x6 transition matrix, $\mathbf{M}$, was produced, where $\mathbf{M}_{i,j}$ was the total number of transitions from AOI i to AOI j . Each row, $\mathbf{M}_{i,*}$, was divided by the sum of row i , so that $\mathbf{M}_{i,*}$ represented the probability of fixation transition from AOI i to any of the six AOIs. Finally, GTE was computed using Equation 2 (Ciuperca and Girardin, 2007), applying the transition matrix $\mathbf{M}$ and the probability vector $\mathbf{V}$.

$$H_{GTE}(M) = - \sum_{i=1}^6 V_i \sum_{j=1}^6 M_{i,j} \cdot \log(M_{i,j})$$

Equation 2

4.2.7 Statistical analysis

Performance measures (i.e., accuracy, planning and execution time) and entropy measures for optimal trials were analyzed using a one-way repeated measures ANOVA with Level (3, 4, 5, 6) as the independent variable. This analysis of performance and entropy measures provides insight into the effect of increasing task difficulty (i.e., cognitive load). Traditional gaze measures were analyzed using a two-way repeated measures ANOVA with AOI (goal space, workspace) and Level (3, 4, 5, 6) as the independent variables. This was specifically done to examine the differences in gaze behaviours across the two AOIs as task difficulty increased. A secondary analysis was completed to compare performance and entropy measures between the optimal and non-optimal trials via a two-way repeated measures ANOVA with Response (optimal, non-optimal) and Level (4, 5, 6) as the independent variables. Level 3 was not included in this analysis because non-optimal trials at this level were rare (<1%). This follow-up analysis was also conducted on traditional gaze behaviours via a 3-way ANOVA with Response (optimal, non-optimal), AOI (goal space, workspace) and Level (4, 5, 6) as the independent variables. All ANOVAs were performed with an alpha level set at 0.05. The Bonferroni post hoc correction for multiple comparisons was applied for all post hoc analyses following the repeated measure ANOVAs to determine significant differences between variables.

4.3 RESULTS

4.3.1 Performance measures

Results for accuracy produced a main effect for level, $F(3, 78)=47.88, p<0.001, \eta_p^2=0.648$. As expected, percentage of optimal trials decreased with each successive increase in task difficulty (i.e., Level 3=97%, SD=8%; Level 4=86%, SD= 7%; Level 5=63%, SD=21%; Level 6=47%, SD=25%). Post-hoc comparisons demonstrated a significant reduction in accuracy from levels 3 to 5 ($p<0.001$), levels 3 to 6 ($p<0.001$), levels 4 to 5 ($p<0.001$), and levels 4 to 6 ($p<0.001$) (**Figure 4.2A**).

The following analysis includes optimal trials only. The results comparing optimal and non-optimal data are presented in section '*Comparison of optimal and non-optimal performance*'. Initial planning time yielded a main effect of level, $F(3, 78)= 24.75, p<0.001, \eta_p^2=0.488$. The time spent in the planning stage was longer as task difficulty increased (i.e., Level 3=3.964sec, SD=1.031sec; Level 4=4.955sec, SD=2.054sec; Level 5=8.309sec, SD=5.402sec; Level 6=10.637sec, SD=4.807sec). Post-hoc comparisons showed significant increases in planning time between levels 3 and 5 ($p=0.002$), levels 3 and 6 ($p<0.001$), levels 4 and 5 ($p=0.009$), and levels 4 and 6 ($p<0.001$) (**Figure 4.2B**). Movement execution time produced a main effect of level, $F(3, 78)=48.09, p<0.001, \eta_p^2=0.649$. The time it took to complete the execution stage increased as task difficulty increased (i.e., Level 3=3.706sec, SD=1.573sec; Level 4=4.823sec, SD=.985sec; Level 5=6.505sec, SD=1.933sec; Level 6=8.245sec, SD=2.108sec). Post-hoc analyses demonstrated significant increases in execution time across all comparisons of increasing task difficulty ($ps<0.018$) (**Figure 4.2C**). These results demonstrate that the task load increased progressively across the levels.

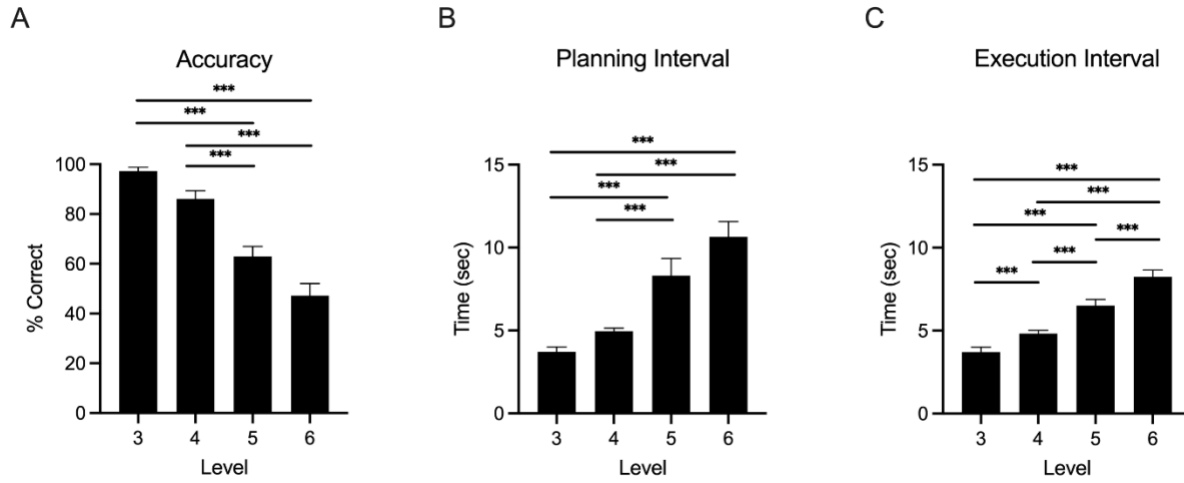


Figure 4.2. Tower of London performance measures. Group mean accuracy scores (%) (A), initial planning time (sec) (B), and movement execution time (sec) (C) are demonstrated across all levels of task difficulty. Only correct trials are reflected in panels (B) and (C). Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

4.3.2 Gaze behaviour during the planning interval

Dwell time (%) demonstrated a main effect of AOI, $F(1, 26) = 12.65$, $p = 0.001$, $\eta_p^2 = 0.327$, level, $F(3, 78) = 7.93$, $p < 0.001$, $\eta_p^2 = 0.234$, and an interaction involving AOI and level, $F(3, 78) = 11.967$, $p < 0.001$, $\eta_p^2 = 0.315$. During the planning interval, dwell time in the workspace (mean=42%, SD=19%) was significantly longer compared to the goal space (mean=31%, SD=10%). Post-hoc comparisons revealed that as task difficulty increased, dwell time in the workspace increased (i.e., Level 3=35%, SD= 18%; Level 4= 38%, SD= 19%; Level 5= 44%, SD= 13%; Level 6=51%, SD=13%) while goal space dwell time decreased (i.e., Level 3=31% SD=6%; Level 4=32%, SD=6%; Level 5=32%, SD=7%; Level 6=26%, SD=8%). Specifically,

significant increases in workspace dwell time were demonstrated between levels 3 and 5 ($p=0.023$), levels 3 and 6 ($p<0.001$), levels 4 and 6 ($p=0.001$), and levels 5 and 6 ($p=0.022$) (**Figure 4.3A**). In contrast, goal space dwell time data revealed a significant decrease between levels 4 and 6 ($p=0.036$), and levels 5 and 6 ($p=0.022$) (**Figure 4.3A**).

Results for fixation rate during the planning interval demonstrated a main effect of AOI, $F(1, 26)=21.20$, $p<0.001$, $\eta_p^2=0.449$, level, $F(3, 78)=7.15$, $p<0.001$, $\eta_p^2=0.216$, and an interaction involving AOI and level, $F(3,78)=15.21$, $p<0.001$, $\eta_p^2=0.369$. Fixation rate was significantly greater in the workspace (mean=1.20 fixations/sec, SD=0.16 fixations/sec), compared to the goal space (mean=0.95 fixations/sec, SD=0.25 fixations/sec). Post-hoc comparisons demonstrate that workspace fixation rate increased with increasing level (i.e., Level 3=0.99 fixations/sec, SD=0.29 fixations/sec; Level 4=1.09 fixations/sec, SD=0.34 fixations/sec; Level 5=1.32 fixations/sec, SD=0.30 fixations/sec; Level 6=1.39 fixations/sec, SD=0.33 fixations/sec), while fixation rate in the goal space decreased (i.e., Level 3=1.05 fixations/sec, SD=0.27 fixations/sec; Level 4=0.97 fixations/sec, SD=0.17 fixations/sec; Level 5=0.99 fixations/sec, SD=0.17 fixations/sec; Level 6=0.81 fixations/sec, SD=0.24 fixations/sec). These respective trends were significant in the goal space between levels 3 and 6 ($p=0.005$), 4 and 6 ($p=0.046$), 5 and 6 ($p=0.003$), and in the workspace between levels 3 and 5 ($p<0.001$), 3 and 6 ($p<0.001$), 4 and 5 ($p=0.001$), and 4 and 6 ($p<0.001$). In terms of fixation duration, no main effects or an interaction were demonstrated (All $F_s<1.02$, $p_s>0.388$).

Results for scan path analysis during the planning stage demonstrated a significant main effect of AOI, $F(1, 26)=19.89$, $p<0.001$, $\eta_p^2=0.433$, level, $F(3, 78)=17.30$, $p<0.001$, $\eta_p^2=0.400$, and an interaction involving AOI and level, $F(3, 78)=11.33$, $p<0.001$, $\eta_p^2=0.303$. Scan path length was significantly longer in the workspace (mean=28.0°, SD=14) compared to the goal

space (mean=20.8°, SD=8). However, post-hoc analyses demonstrated that as task difficulty increased there was a greater increase in scan path length within the workspace (i.e., Level 3=14.6°, SD=5.2°; Level 4=19.3°, SD=9.3°; Level 5=35.2°, SD=27.8°, Level 6=43.0°, SD=26.1°) compared to the goal space (i.e., Level 3=14.5°, SD=7.1°; Level 4=16.5°, SD=7.4°; Level 5=25.3°, SD=14.5°; Level 6=27.3°, SD=14.7°). This became significantly different at levels 4, 5, and 6 ($p < 0.015$) (**Figure 4.3C**).

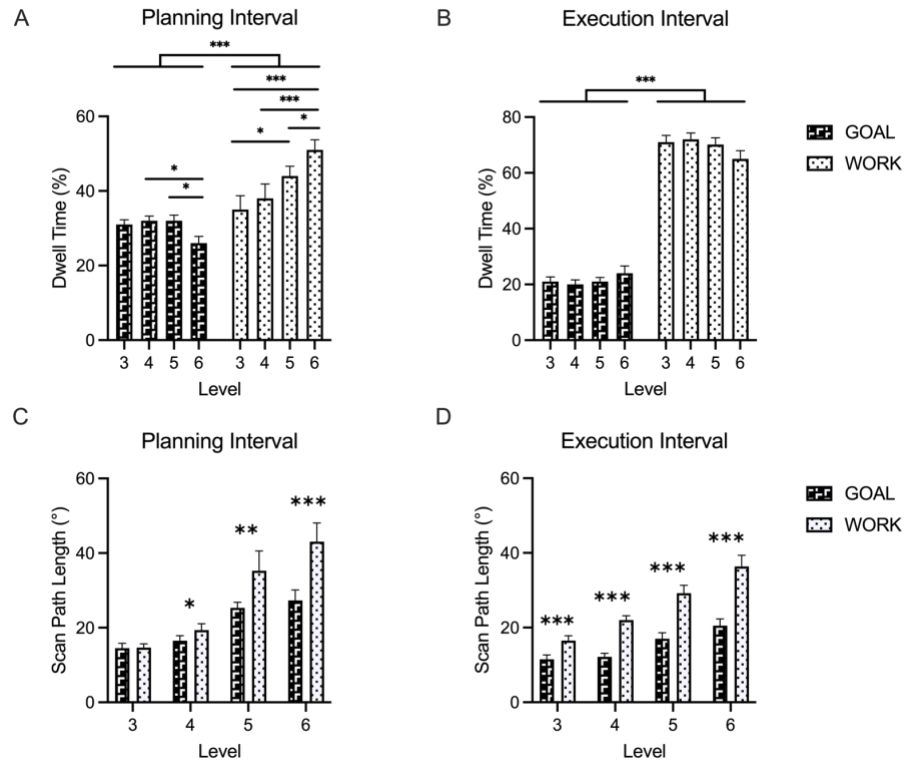


Figure 4.3. Group mean dwell time (%) during the planning interval (A) and execution (B) and scan path length (°) during the planning interval (C) and execution (D) are demonstrated across all levels of task difficulty. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Further analysis of the gaze data was conducted to examine whether the location of a participant's first fixation influenced any of the reported performance or gaze measures. The data were segregated based on the AOI where participants fixated first (i.e., goal vs. workspace). A chi-square analysis demonstrated that more trials began with participants looking toward the workspace first ($n=415$; 68%) compared to the goal space ($n=191$; 32%), $X^2(1)=82.8$, $p<0.001$. Data were subjected to a 2 (AOI: goal space, workspace) by 4 (Level: 3, 4, 5, 6) repeated measures ANOVA. Initial planning time was trending toward significance for AOI, $F(1, 26)=3.04$, $p=0.073$, $\eta_p^2=0.105$. Specifically, longer planning times were more likely to be observed when participants fixated in the goal space first (mean=7.191sec, SD= 2.280sec) compared to when they fixated in the workspace first (mean=6.464ms, SD= 2.005sec) (**Figure 4.4**). No interaction involving AOI and level was observed. Note that a main effect of level was also significant, but this is reported above. First fixation location was not associated with any other differences across performance or gaze behaviours examined in the study.

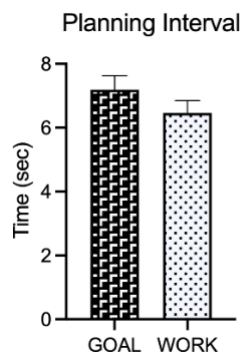


Figure 4.4. Group mean initial planning interval (sec) when participants looked in the goal space versus the workspace first. Error bars represent SEM.

4.3.3 Gaze behaviour during the execution interval

Dwell time (%) during the execution interval revealed a main effect of AOI, $F(1, 26)=360.1$, $p<0.001$, $\eta_p^2=0.933$. Workspace dwell time (mean=69%, SD=14%) was significantly longer than goal space dwell time (mean=21%, SD=10%) during solution execution. No main effect of level or higher-level interactions in either AOI were demonstrated (**Figure 4.3B**). In terms of fixation rate during the execution interval, data demonstrated a main effect of AOI, $F(1, 26)=247.512$, $p<0.001$, $\eta_p^2=0.905$. Specifically, fixation rate in the goal space was significantly lower (mean=0.69 fixations/sec, SD=0.14 fixations/sec) compared to the workspace (mean=1.65 fixations/sec, SD=0.27 fixations/sec). Similarly, fixation duration during the execution interval demonstrated a main effect of AOI, $F(1, 26)=132.4$, $p<0.001$, $\eta_p^2=0.836$. Goal space fixation duration was shorter (mean=318ms, SD=60ms) than workspace fixation duration (mean=439ms, SD=58ms). No other main effects or interactions were revealed in fixation rate or fixation duration during the execution interval.

Results for scan path analysis during the execution stage demonstrated a significant main effect of AOI, $F(1, 26)=98.96$, $p<0.001$, $\eta_p^2=0.792$, level, $F(3, 78)=23.37$, $p<0.001$, $\eta_p^2=0.473$, and an interaction between AOI and level $F(3, 78)=8.52$, $p<0.001$, $\eta_p^2=0.247$. Similar to the planning stage, scan path length during the execution stage was significantly longer in the workspace (mean=26°, SD=7.2°) compared to the goal space (mean=15°, SD=4.9°). Scan path length increased within both AOIs as task difficulty increased. However, post-hoc comparisons demonstrated that scan path length increased to a much larger degree with every subsequent level of task difficulty within the workspace (i.e., Level 3= 16.5°, SD= 7.1°; Level 4=22°, SD= 6.4°; Level 5=29.2°, SD=10.9°; Level 6=36.4°, SD=15.7°) (all $ps<0.035$) compared to the goal space

(i.e., Level 3= 11.5°, SD=6.3°; Level 4=12.2°, SD=4.8°; Level 5=17°, SD=8.6°; Level 6=20.5°, SD=9.7°) (all $ps < 0.578$) (**Figure 4.3D**).

4.3.4 The influence of task difficulty on gaze entropy

During optimal task performance, SGE during the planning interval was the only measure shown to violate the Shapiro-Wilk test for normality ($df=27$, $t= 0.901$, $p=0.014$). Accordingly, a Friedman's test was conducted and demonstrated a significant main effect of level ($df=3$, $t=11.311$, $p=0.010$). Pairwise comparisons showed that the main effect was driven by a significant increase in planning SGE limited to levels 4 and 5 ($p=0.013$). In contrast, SGE revealed no significant main effect of level in the execution interval ($Fs < 0.855$, $ps > 0.468$). Therefore, the dispersion of fixations is relatively similar across all difficulty levels during planning (**Figure 4.5A**) and solution execution (**Figure 4.5B**). In contrast, GTE yielded a main effect of level during the planning interval, $F(3, 78)=16.776$, $p < 0.001$, $\eta_p^2=0.392$, as well as during the execution interval, $F(3, 78)=41.876$, $p < 0.001$, $\eta_p^2=0.617$. Specifically, data revealed an increase in the transitional entropy of gaze patterns as task difficulty increased in the planning (i.e., Level 3= 0.92 bits, SD= 0.23 bits; Level 4=1.04 bits, SD=0.32 bits; Level 5=1.33 bits, SD=0.23 bits; Level 6=1.29 bits, SD=0.38 bits) (**Figure 4.5C**) and execution (i.e., Level 3= 0.94 bits, SD= 0.19 bits; Level 4=1.11 bits, SD=0.18 bits; Level 5=1.25 bits, SD=0.19 bits; Level 6=1.36 bits, SD=0.14 bits) (**Figure 4.5D**) intervals. Post hoc comparisons demonstrated significant increases in GTE during planning between levels 3 and 5 ($p < 0.001$), levels 3 and 6 ($p=0.001$), levels 4 and 5 ($p < 0.001$), and levels 4 and 6 ($p=0.031$). Significant increases in GTE during solution execution were also found between levels 3 and 4 ($p=0.002$), levels 3 and 5

($p < 0.001$), levels 3 and 6 ($p < 0.001$), levels 4 and 5 ($p = 0.018$), and levels 4 and 6 ($p < 0.001$). All normalized entropy values are reported in **Table 4.1** using the H_{max} value of 2.585.

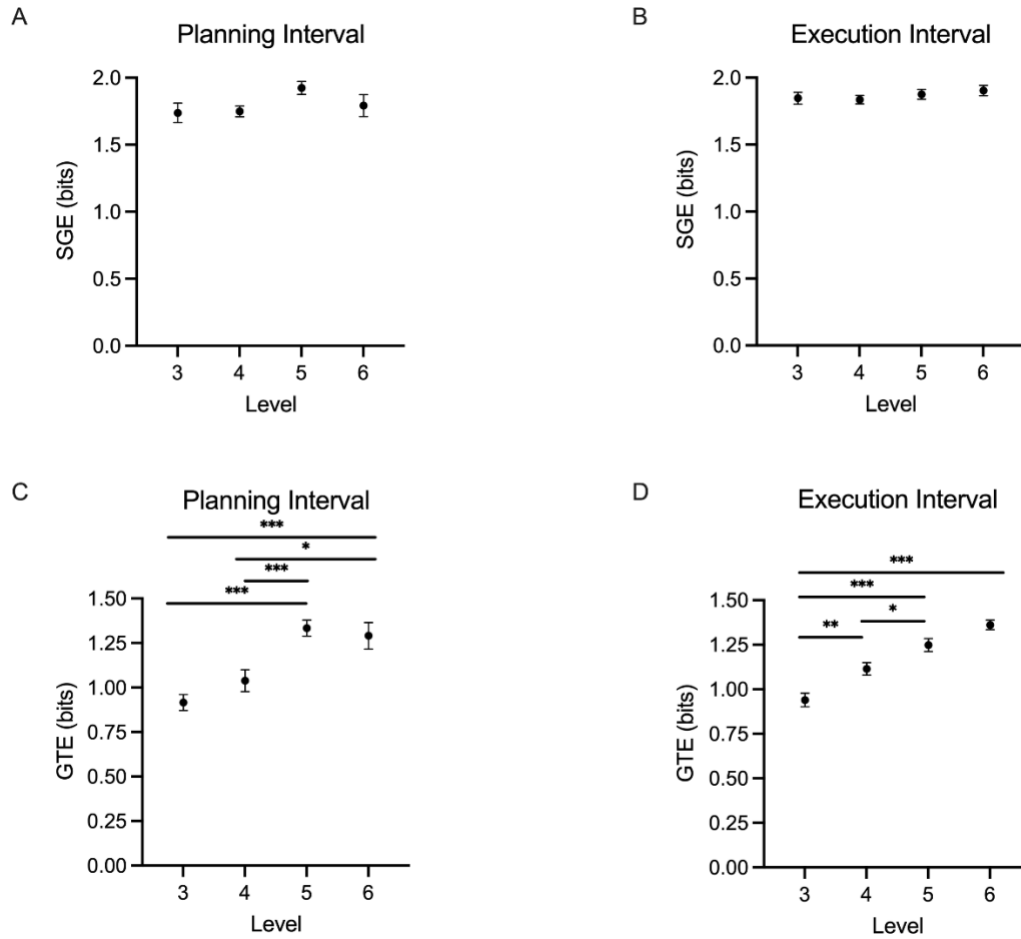


Figure 4.5. Group mean entropy (bits) values for stationary gaze entropy (SGE) (A and B) and gaze transition entropy (GTE) (C and D) during the planning and execution intervals, respectively, across all levels of task difficulty. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

<i>OPTIMAL</i>	<i>SGE</i>		<i>GTE</i>	
<i>Level</i>	Planning	Execution	Planning	Execution
3	0.672(0.15)	0.715(0.09)	0.354(0.09)	0.364(0.08)
4	0.676(0.08)	0.710(0.07)	0.402(0.12)	0.431(0.07)
5	0.744(0.10)	0.726(0.07)	0.516(0.09)	0.483(0.07)
6	0.6932(0.16)	0.736(0.08)	0.499(0.14)	0.526(0.05)

Table 4.1. Normalized group mean entropy values for stationary gaze entropy (SGE) and gaze transition entropy (GTE) during the planning and execution intervals, respectively, across all levels of task difficulty. Standard deviation values reported in brackets.

4.3.5 Comparison of optimal and non-optimal performance

Additional analyses were conducted to examine differences between the non-optimal and optimal trials. As such, we only reported significant effects and interactions involving response (i.e., non-optimal versus optimal trials). Non-optimal trials rarely occurred at level 3 (<1% of trials). Therefore, the analysis excluded all level 3 trials.

Behavioural results demonstrated a difference between optimal and non-optimal trials. Initial planning time was not statistically different between the optimal and non-optimal trials (**Figure 6A**). In contrast, movement execution time yielded a main effect of response, $F(1, 26)=306.22$, $p<0.001$, $\eta_p^2=0.922$. The duration of the execution stage was significantly prolonged during non-optimal trials (mean=19.501sec, SD=3.998sec) compared to their optimal counterparts (mean=6.527sec, SD=1.142sec) (**Figure 6B**). No interaction involving level and response were revealed for movement execution time.

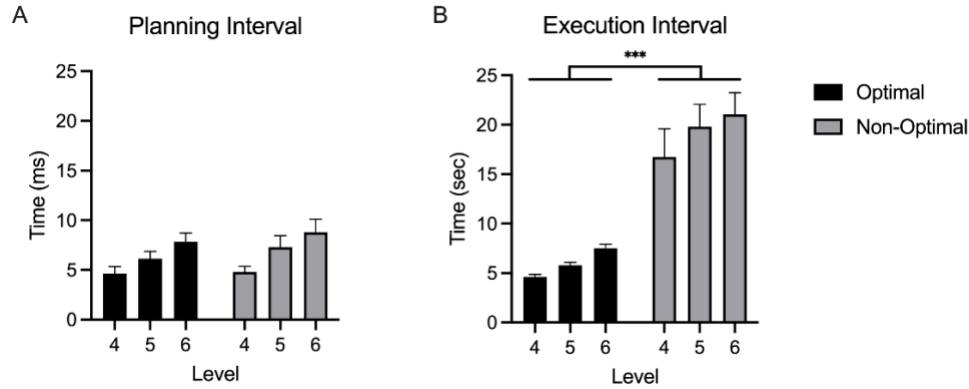


Figure 4.6. Group mean initial plan time (sec) (A) and movement execution time (sec) (B) are illustrated for optimal (black) and non-optimal (grey) data as a function of increasing task difficulty. Error bars represent SEM. *** $p \leq 0.001$.

With respect to dwell time, results did not reveal any main effects or interactions involving response during the planning interval, $F_s(2, 52) < 1.6$, $p_s > 0.230$. However, during the execution interval dwell time demonstrated a main effect of response, $F(1, 26) = 7.73$, $p = 0.010$, $\eta_p^2 = 0.229$, as well as a two-way interaction between response and AOI, $F(1, 26) = 9.783$, $p = 0.004$, $\eta_p^2 = 0.273$. The two-way interaction demonstrated that compared to optimal trials, non-optimal trials were associated with longer dwell times in the goal space (mean=21%, SD =6% vs. 24%, SD=9%) and shorter dwell times in the workspace (mean=69%, SD =6% vs. 62%, SD=9%) (all $p_s < 0.05$) during solution execution.

Fixation rate in the planning interval demonstrated a main effect of response, $F(1, 26) = 8.20$, $p = 0.008$, $\eta_p^2 = 0.240$. Non-optimal trials were associated with lower fixation rates (mean=2.28 fixations/sec, SD=0.31 fixations/sec) compared to their optimal counterparts (mean=2.40 fixations/sec, SD=0.28 fixations/sec) (**Figure 4.7A**). Fixation duration during the planning stage also demonstrated a main effect of response, $F(1, 26) = 13.93$, $p = 0.001$, $\eta_p^2 = 0.349$.

Non-optimal trials were associated with longer fixation durations (mean=434ms, SD=85ms) compared to their optimal counterparts (mean=389ms, SD=53ms) (**Figure 4.7B**). No other interactions involving response were demonstrated for fixation rate or duration during the planning stage. In contrast, fixation rate and fixation duration during solution execution did not demonstrate any main effects or higher-level interactions involving response (all $F_s(1, 26)=0.023$ and 2.047 , $p_s>0.164$). The one exception to this was fixation duration, which demonstrated an interaction between response and AOI, $F(1, 26)=13.02$, $p=0.001$, $\eta_p^2=0.334$. The two-way interaction demonstrated that in comparison to optimal trials, non-optimal trials were associated with longer goal space fixation durations (mean=320ms, SD= 58ms vs. mean=390ms, SD=121ms) and shorter workspace fixation durations (mean=438ms, SD=61ms vs. mean=400ms, SD=61ms) during solution execution.

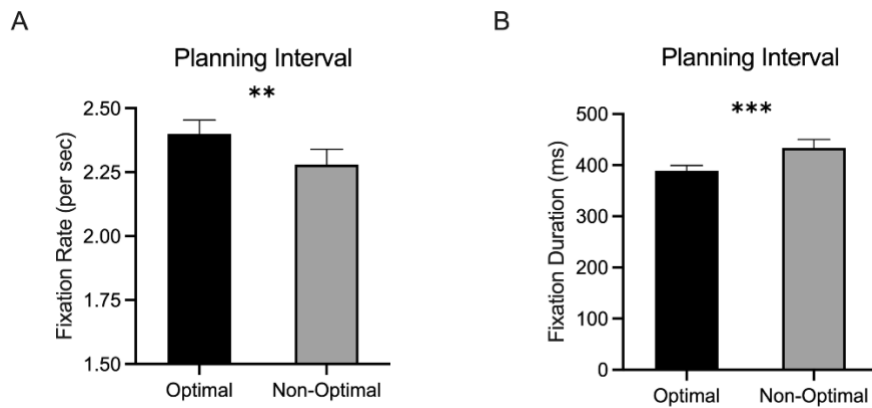


Figure 4.7. Group mean fixation rate (fixations per sec) (A) and fixation duration (ms) (B) during the planning interval are illustrated for optimal (black) and non-optimal (grey) data. Error bars represent SEM. * $p\leq0.05$, ** $p\leq0.01$, *** $p\leq0.001$.

Lastly, scan path length during the planning stage did not reveal any main effect or interactions involving response, all $F_s(2,52)>0.61$, $p_s>0.43$. In contrast, scan path length during the execution interval presented a main effect of response, $F(1, 26)=177.15$, $p<0.001$, $\eta_p^2=0.872$, a significant interaction between response and AOI, $F(1, 26)=28.60$, $p<0.001$, $\eta_p^2=0.524$, and a three-way interaction between response, AOI and level, $F(2, 52)=12.23$, $p<0.001$, $\eta_p^2=0.320$. Participants exhibited significantly longer scan paths during non-optimal trials (mean=60°, SD=16°) compared to optimal trials (mean=23°, SD=6°). Post-hoc comparisons examining the three-way interaction demonstrated that optimal trials demonstrated significant increases in scan path length during movement execution, particularly within the workspace ($p_s<0.035$), as level increased (**Figure 4.8A**). A similar trend was demonstrated within non-optimal trials; however, the observed trend of increasing scan path length as a function of task difficulty seemed to plateau between levels 5 and 6 within the workspace ($p=0.275$) for non-optimal trials (**Figure 4.8B**).

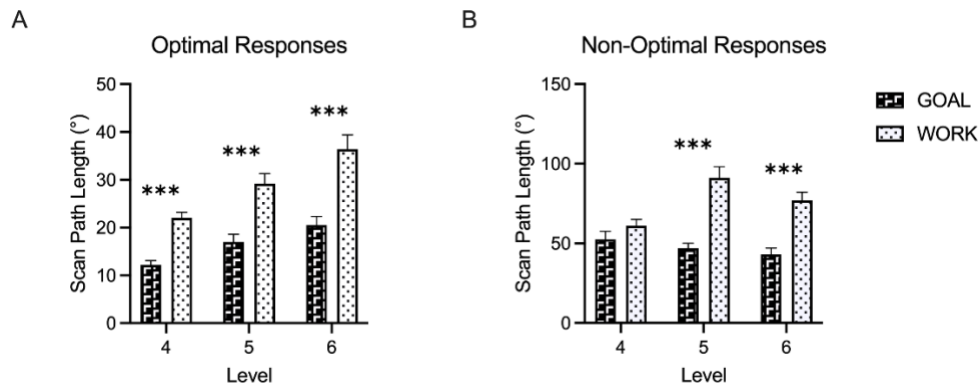


Figure 4.8. Group mean scan path length (°) is demonstrated across all levels of task difficulty for optimal (A) and non-optimal (B) responses during the execution interval. Error bars represent SEM. * $p\leq0.05$, ** $p\leq0.01$, *** $p\leq0.001$.

SGE demonstrated a main effect of response during the planning interval, $F(1, 26)=23.754$, $p<0.001$, $\eta_p^2=0.477$. Specifically, the dispersion of fixations was significantly larger during non-optimal performance (mean= 2.03 bits, SD=0.13 bits) compared to their optimal counterparts (mean=1.82 bits, SD=0.21 bits) (**Figure 4.9A**). Notably, no main effect of response was demonstrated in the execution interval ($F=0.371$, $p=0.548$) (**Figure 4.9B**).

GTE yielded a main effect of response during the planning interval, $F(1, 26)=55.379$, $p<0.001$, $\eta_p^2=0.681$, as well as an interaction between response and level, $F(2, 52)=3.424$, $p=0.04$, $\eta_p^2=0.116$. **Figure 4.9C** demonstrates that GTE was significantly larger during non-optimal trials (mean=1.60 bits, SD= 0.15 bits) compared to their optimal counterparts (mean=1.22 bits, SD= 0.24 bits) during the planning interval. Post hoc analysis demonstrated that the interaction was due to the increase in GTE as task difficulty increased in optimal trials, specifically between levels 4 and 5 ($p=0.018$). Notably, this was not evident across non-optimal trials since GTE was higher, but consistent across all levels of difficulty. No main effect of response was revealed in GTE during the execution stage ($F=0.524$, $p=0.475$) (**Figure 4.9D**). All normalized entropy values are reported in **Table 4.1** using the H_{max} value of 2.585.

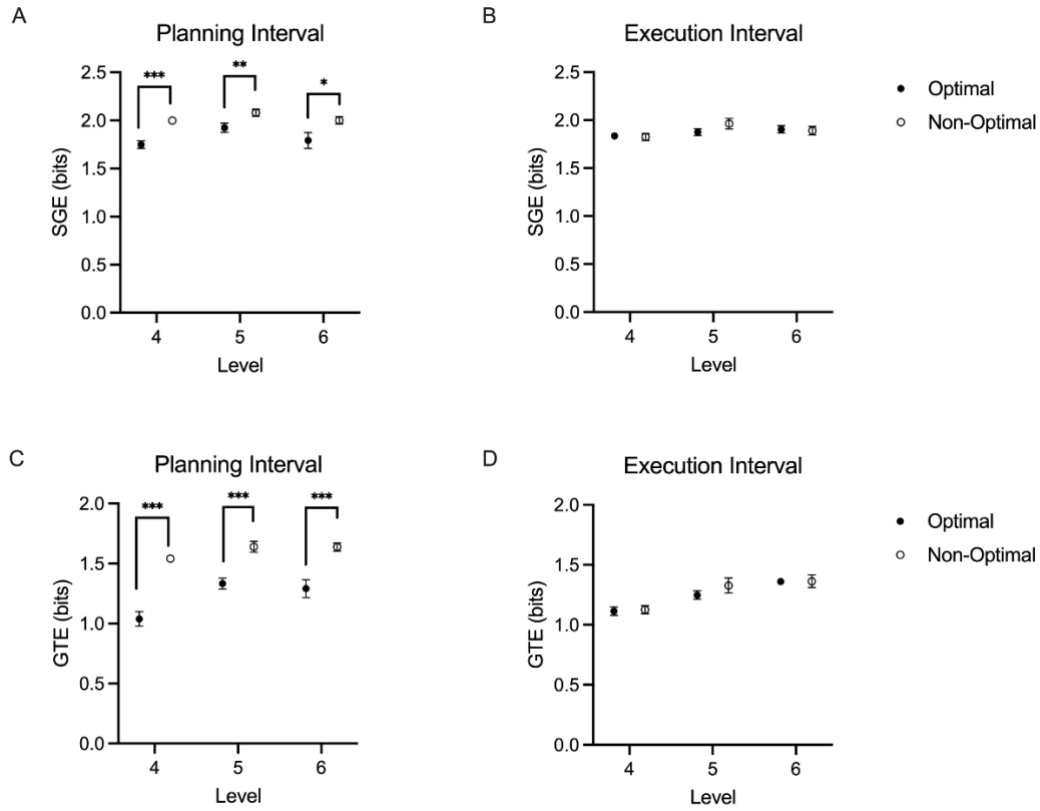


Figure 4.9. Group mean entropy (bits) values are shown for optimal (filled black circles) and non-optimal (open circles) stationary gaze entropy (SGE) (A and B) and gaze transition entropy (C and D) across task difficulty levels 4-6 during the planning and execution intervals, respectively. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

NON- OPTIMAL	SGE		GTE	
Level	Planning	Execution	Planning	Execution
4	0.773(0.05)	0.706(0.07)	0.596(0.06)	0.436(0.06)
5	0.805(0.08)	0.760(0.11)	0.634(0.09)	0.514(0.12)
6	0.775(0.08)	0.731(0.9)	0.633(0.07)	0.527(0.11)

Table 4.2. Normalized group mean entropy values for non-optimal stationary gaze entropy (SGE) and gaze transition entropy (GTE) during the planning and execution intervals, respectively, across task difficulty levels 4-6. Standard deviation values reported in brackets.

4.4 DISCUSSION

This study set out to systematically examine the utility of using gaze-based metrics to objectively characterize the performance proficiency of healthy adults during a complex planning and problem-solving task that progressively increased in difficulty. Additionally, an in-depth analysis was conducted on non-optimal trials to determine whether putative changes in gaze behaviour could identify the erroneous planning and problem-solving processes that lead to non-optimal performance. In support of our hypotheses, increasing task difficulty was associated with a workspace-specific increase in dwell time which was restricted to the planning interval. Though the execution interval did not show a similar effect of task difficulty, there was a general dwell time-bias towards the workspace compared to the goal space. Results indicate that entropy metrics (i.e., SGE and GTE) provide additional information pertaining to the complexity (i.e., dispersion and unpredictability) of gaze behaviour. Specifically, GTE (i.e., gaze unpredictability) significantly increased as a function of task difficulty (i.e., cognitive load) during the initial

planning and execution intervals. Furthermore, non-optimal trials were associated with a reduction in fixation rate, an increase in fixation duration, and an increase in SGE and GTE during the planning stage of non-optimal trials. Taken together, these results provide clear evidence for how entropy metrics can be employed in conjunction with traditional gaze analyses to 1) objectively identify increases in cognitive load (related to the increasing demands on planning and visuospatial working memory) and 2) distinguish putative changes in gaze behaviour within the early stages of planning that are indicative of erroneous planning and problem-solving processes associated with poor performance. In outlining our results, we first describe the performance results for optimal and non-optimal trials. Subsequently, we describe how the strategic distribution of gaze within the planning and execution intervals contributes to the discrete cognitive processes and strategies that underlie optimal task performance. Lastly, we describe how they differ from their non-optimal response counterparts.

4.4.1 Performance indicators of increasing task difficulty and non-optimal performance

Increasing task difficulty (i.e., the minimum number of moves required to solve the problem) was associated with a reduction in accuracy and an increase in initial planning and movement execution times. In line with previous work (Hodgson et al., 2000; 2002; Huddy et al., 2007), these behavioural findings confirm that as the problems become more difficult there is a progressive increase in the time taken to plan and execute the correct sequence of ball movements to achieve the task goal. There was also an increase in the number of erroneous moves during more difficult problems (Hodgson et al., 2000; 2002; Huddy et al., 2007). Indeed, there were performance differences between optimal and non-optimal trials. However, it is important to note that the resolution of these performance measures did not allow for the

detection of early indicators of non-optimal behaviour during the planning interval. Instead, our results demonstrated that non-optimal movement execution times were approximately 13 seconds longer than optimal trials. A finding taken to reflect the longer time required to reverse erroneous movements made during the execution interval and re-sequence sub-optimal solutions that were initially strategized during the planning interval.

4.4.2 Gaze characteristics during planning

Seminal work involving the analysis of gaze behaviour has reliably demonstrated how several spatial and temporal gaze characteristics reflect the various cognitive processes that facilitate planning and problem-solving in the TOL task (Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012). With respect to the planning stage, there is clear evidence of a selective gaze bias toward the workspace. This indicates that attention is overtly deployed to that region (Hodgson et al., 2000; 2002; Kaller et al., 2009; Nitschke et al., 2012). In the current study, relative dwell time, fixation rate and scan path length were significantly higher/longer in the workspace compared to the goal space, and they all displayed a workspace-specific increase as task difficulty increased. Analogous with previous studies (Hodgson et al., 2000; 2002; Kaller et al., 2009; Nitschke et al., 2012; Tremblay et al., 2006), our results suggest that a greater amount of time and attention is devoted to this area of interest for solution elaboration and to perform/maintain mental manipulations of the visuospatial arrays. The current study further extends this research by examining gaze entropy, which objectively characterizes the complexity (i.e., spatial distribution and sequence predictability) of scanning behaviour. First, our results reveal that SGE did not change as task difficulty increased. Though this is in line with previous suggestions that SGE is a proxy of perceptual demand (Shiferaw et al., 2018), it is interesting to

note that this measure of fixation dispersion did not capture the workspace bias effects that were apparent in the traditional gaze measures. In contrast, planning stage GTE increased as task difficulty increased. As expected, planning stage GTE further supports the traditional gaze findings that show an increase in the complexity of gaze behaviours associated with an increase in task difficulty. Specifically, the increase in GTE indicates that the sequence of fixations included more transitions between different AOIs in a more unpredictable and less structured pattern. The current findings show clear evidence that as task difficulty increased, the observed changes in gaze behaviours are indicative of the increasing complexity associated with solution elaboration, mental manipulations of the visuospatial arrays, and the increased visuospatial memory demands involved in extracting information during the initial planning interval of more difficult problem sets.

A novel finding in the current study was the observation of a decrease in relative dwell time spent in the goal space at the highest level of task difficulty (i.e., level 6). Notably, goal space dwell time has been suggested to reflect the time it takes to generate an internal representation of the goal state and assess problem goals (Hodgson et al., 2000; Kaller et al., 2009). Previous work has reported no differences in goal space dwell time across lower levels of the TOL task during the planning stage (Hodgson et al., 2000; 2002). In accounting for this difference, it is important to note that the current study reports dwell time as a relative percentage to account for the increase in time spent in the planning stage across difficulty levels. Therefore, a plateau in the actual amount of time spent in the goal space at these higher levels coupled with an increase in initial planning time produced a reduction in the goal space dwell time percentage. This is also demonstrated in the fixation rate observed in the goal space, which decreases with increasing task complexity. The reduction in goal space dwell time and fixation

rates at these higher levels are proposed to be a result of the limitations associated with visuospatial working memory capacity (Ballard et al., 1997; Baddeley, 2003). In favor of devoting more time and attentional resources toward the workspace for solution elaboration and maintenance, working memory constraints produce a trade-off effect resulting in a reduction in the relative time and attention devoted toward the goal space during the planning interval. Contrary to the decreases in fixation rate and dwell time observed in the goal space during the planning stage, scan path length increased as task complexity increased; though, the goal space effect was much less robust compared to the workspace. It is important to note that scan path length is more indicative of the spatial distribution of eye movement patterns (i.e., a broad measure of dispersion and complexity), rather than the temporal sequence of gaze. We also reported an increase in GTE, which suggests that the pattern of eye movements employed at these higher levels was more complex during the initial planning phase of performance. Taken together, gaze analysis provides evidence that eye movement scan pattern becomes more complex as task difficulty increases and demonstrate specific gaze strategies via regional biases that aid in the planning of longer and more complex solution sequences.

An important consideration of gaze analysis during cognitive task performance is the involvement of specific gaze strategies that emerge at the beginning of the task. For instance, Martin et al. (2017) concluded that maintaining a covert gaze strategy (i.e., fixation remains directed to a central target) during the encoding phase of the Corsi-block task was associated with improved performance compared to an overt gaze strategy (i.e., fixations follow the sequence of targets). Further, Kaller et al. (2009) reported that the total number of gaze shifts during TOL planning was significantly associated with the directional preference of an individual's first fixation (i.e., goal space vs. workspace). Expanding on this, the current findings

provide preliminary evidence of an association between total planning time efficiency and the region of an individual's first fixation (i.e., ~727ms increase in planning time when eyes were first directed towards the goal space vs. workspace). Recall that eight trials were completed at TOL levels 4, 5, and 6. Therefore, only a few trials across these conditions were able to be analyzed once the data were segregated further to account for where an individual's first fixation occurred. As such, more research is required to determine if 'where we look first' plays a role in the efficiency of completing several sub-components of planning including comparing and matching individual features of the visuospatial configuration, identifying differences, generating an internal representation of the goal state, mentally manipulating visuospatial arrays, and solution elaboration and maintenance.

4.4.3 Gaze characteristics during solution execution

A detailed analysis of the spatial and temporal gaze characteristics during the planning stage of the TOL task has been reported in previous work (Hodgson et al., 2000; 2002; Huddy et al., Kaller et al., 2009; Nitschke et al., 2012). However, little to no analysis of the execution stage has been completed to date. Kaller et al. (2009) characterized the dwell time and the distribution of gaze shifts specific to the workspace in TOL problem sets involving 3 ball moves or less. The current findings characterized relative dwell time, fixation rate, fixation duration, scan path length, SGE and GTE across more complex problem sets. Our findings provide evidence of a change in the gaze strategies from those employed in the planning stage in order to support the different goals that are involved in the execution stage. Specifically, a substantial gaze bias towards the workspace was identified in the dwell time, fixation rate, fixation duration, and scan path length measures during the execution interval. Indeed, the time and attention

devoted to this region during the execution period support the fact that movement processes and solution execution occur almost exclusively in the workspace (Kaller et al., 2009). Though there are times at which attention is re-focused into the goal space, this is mainly completed to re-fixate the goal state and verify task completion (Kaller et al., 2009). This is also supported by the finding that fixation duration was shorter in the goal space compared to the workspace. The process of re-verifying information in the goal space which is being held in working memory takes less time (i.e., shorter fixation durations) than problem solving in the workspace. The results clearly show that solution execution processes (i.e., identifying, matching, recalling and adjusting pieces of information relevant and necessary to the planned sequence of ball movements) are predominantly supported through eye movements in the workspace, which are characterized by longer fixations. Moreover, the longer fixation times are also crucial to supporting the motor requirements of the task during the execution stage (i.e., fixating the target and moving the cursor to select the ball, then fixating the new location and moving the cursor to drop the ball). Last, scan path length and GTE were the only measures to demonstrate an increase associated with task difficulty. The current findings suggest that the increase in scan path length during more difficult problem sets is linked to an increase in GTE. More specifically, as task difficulty increased the gaze pattern became more complex due to more frequent fixation transitions between AOIs in an increasingly random sequence. Notably, this was seen together with no change in SGE, which suggests that the relative distribution of fixation locations remained the same despite an increase in task difficulty. Indeed, no other gaze characteristics have been shown to be modulated by task difficulty at this stage that could potentially shift the distribution of fixations within the task environment. Therefore, our findings suggest that the longer and increasingly complex solution sequences required to complete more difficult problem

sets are associated with more random, unstructured- and thus, longer- fixation patterns that occupy a similar spatial dispersion pattern relative to those seen in the solution execution stage of easier problem sets.

4.4.4 Gaze characteristics distinguish optimal from non-optimal performance

The broader significance of the present findings lies in its application to the identification and execution of near-optimal gaze sequences that are crucial for many day-to-day tasks, and even more so in highly complex occupational environments (i.e., aviation, police, medicine, etc.). Beyond the explicit indications of problem-solving impairments seen through increased error rates, planning and movement times, the analysis of eye movements may offer additional insights into cognitive and general visuomotor processes underlying task performance. Indeed, a growing number of studies examining gaze behaviour and TOL performance in healthy and neurologically impaired populations have asserted that the analysis of eye movements provides a reliable method to identify different stages of the problem-solving process (Hodgson et al., 2000; 2002; Huddy, 2007; Kaller et al., 2009; Nitschke et al., 2012). Moreover, previous work proposed that eye movement analysis has the potential to facilitate a more discriminant attribution of impairments in problem-solving to specific errors in distinct cognitive phases of the overall process (Hodgson et al., 2002; Huddy et al., 2007; Kaller et al., 2009). Specifically, work by Hodgson et al. (2002) reported that Parkinson's patients divided their attention equally between the goal space and the workspace. Thus, demonstrating how planning deficits associated with Parkinson's Disease (PD) are tied to abnormal encoding and maintenance of task-relevant information. This was suggested to reflect abnormalities in the fronto-striatal circuits implicated in working memory and attentional control in PD. In a similar study, Huddy and colleagues

(2007) found that individuals with first episode schizophrenia or schizoaffective disorder demonstrated intact gaze biases and used the same cognitive strategies seen in healthy controls when completing the TOL task. However, these patients showed increased durations of long-gaze periods toward the goal and workspace. The authors suggested that this specific gaze characteristic seen in patients was tied to difficulties in encoding essential features of the stimulus array; a finding taken to evince the slowing of working memory consolidation. Taken together, these studies show how eye movement analysis can identify the source of planning abnormalities among different neurological disorders in a way that more superficial behavioural analysis could not. It is important to note that this could be further explored to examine how gaze behaviour changes across different neurological populations during TOL task completion as a means to gauge impairments in working memory, cognitive flexibility, inhibitory control, or other nonspecific impairments (i.e., distractibility, low motivation) that impact performance (Hodgson et al., 2002; Huddy et al., 2007). However, what is more germane to the current study is the lack of work examining the source of erroneous planning processes that resulted in non-optimal responses. The need to explore this question comes from research demonstrating the potential of using eye-tracking measures as a method to predict upcoming errors in task performance (Ayala and Niechwiej-Szwedo, 2021). The present study specifically sought to employ gaze analysis to obtain a better and more detailed understanding of the eye movement patterns involved in erroneous planning behaviour in healthy adults.

With respect to the planning stage, several crucial differences in gaze behaviour were seen between optimal and non-optimal responses. Despite the equivocal dwell time and scan path length results, fixation rate decreased, and fixation duration increased across both areas of interest during non-optimal compared to optimal trials. Though an increase in fixation duration

provides more time for information extraction, the observed reduction in fixation rate points to two potential sources of early planning errors that underlie the generation of non-optimal responses. First, a reduction in fixation rate suggests that less scanning of pertinent task-related information was completed during these non-optimal trials, which may be insufficient for the generation of an internal representation of the goal state and solution elaboration (Hodgson et al., 2000; Kaller et al., 2009). Second, fixation count is typically associated with the maintenance of information in visuospatial working memory (Clark, 1999; Tremblay et al., 2006). Therefore, the reduction in fixation rate seen in non-optimal trials may lead to the deterioration of the solution sequences that are being generated and maintained in visuospatial working memory. This is further supported by the significantly higher SGE and GTE during the planning stage, which were observed in the non-optimal trials compared to their optimal counterparts. Specifically, non-optimal performance was associated with a wider dispersion of fixations that included more transitions between AOIs in an increasingly random order. Notably, this finding also demonstrates that SGE may in fact be a proxy for something more than stimulus properties. Given that scanning occurred in a less structured manner across a more dispersed area of exploration, our findings evince that the gaze behaviour during non-optimal trials is more characteristic of visual exploration rather than task-engagement (Shiferaw et al., 2018). Due to the complexity of our paradigm, we were unable to directly replicate the analyses used by Hodgson et al. (2000) to examine the gaze behaviour of task-efficient planners. They revealed that gaze was effectively focused on problem-critical balls. However, our combined entropy and traditional gaze analyses provide some support for how the distribution of gaze was more structured (i.e., more concentrated) and likely focused on problem-critical areas necessary for success during optimal task performance. Evidently, the non-optimal gaze patterns observed here

likely underlie, 1) errors in the generation of an accurate internal representation of the goal state, 2) ineffective/disrupted solution elaboration processes, and 3) a disruption in the maintenance of information in visuospatial working memory; thus, interfering the planning processes that are required for optimal task performance. Together, traditional and entropy-based gaze analyses demonstrate how gaze becomes inefficient during the early stages of planning and, more importantly, provide an objective and quantifiable gaze-based marker of non-optimal performance.

Gaze analysis of the solution execution stage continued to demonstrate a workspace selective gaze bias in both optimal and non-optimal responses. However, gaze pattern differences that were identified in the current study during non-optimal trials indicate that the processes normally completed during the initial planning stage were-to some extent- repeated in the execution stage. Specifically, non-optimal trials were associated with an increase in dwell time, fixation duration and scan path length in the goal space, relative to their optimal response counterparts. This was seen in parallel with a reduction in dwell time and fixation duration in the workspace. These results suggest that non-optimal trials were associated with more attention being focused toward the goal space- at the expense of the workspace- during the execution stage to support two corrective actions: 1) the regeneration of an internal representation of the goal state and, 2) to compare the goal space to the current state of erroneous ball movements made in the workspace and adjust the planned solution sequence accordingly. This is further supported by the increase in scan path length found in the execution stage of non-optimal trials across both areas of interest. Therefore, gaze behaviour analysis demonstrates a reliance on increased attention (i.e., dwell time and fixation rate) toward the goal space and increased movement execution time to complete additional planning processes (i.e., reversing erroneous movements,

re-sequencing sub-optimal solutions, reassessing the problem goal) during the execution stage of non-optimal trials to achieve the task goal, albeit using a greater number of moves than necessary. Interestingly, these differences were not associated with any significant changes in SGE or GTE. Instead, our results indicated that the distribution and predictability of gaze during the solution execution stage of non-optimal responses are relatively in line with that of optimal responses. Accordingly, the current results suggest that entropy measures alone do not provide a comprehensive account of gaze behaviour and should therefore be used in conjunction with additional behavioural and gaze-based measures in addressing neurocognitive performance.

4.4.5 Limitations

The results are constrained by at least four methodological limitations. First, the small number of trials used here limited the potential to examine some deeper questions associated with factors that extend beyond the question of task difficulty and response type (i.e., optimal vs. non-optimal). As such, future studies should use more extensive paradigms of the TOL task that allow for more than eight trials to be completed at all levels of difficulty. Second, the current study proposed a novel approach to the study of planning and problem-solving performance via gaze analysis of preferred eye movement patterns/strategies. We acknowledge, however, that this approach is essentially descriptive, and the gap between optimal eye movement patterns/strategies and neurocognitive performance requires additional research. Third, our participant sample consisted of a convenience sample of university students and was therefore not representative of the general population in terms of age and education. Last, the methodological challenges imposed by the Vienna testing environment limited real-time synchronization between coordinated eye and hand movements. Further research examining the

temporal sequence of fixations and its timing relative to hand/mouse movements used to manipulate the balls in the TOL environment are needed to address this issue. For instance, it would be crucial to address the basic relationship and temporal contingency between specific eye movements and motor movements to determine whether specific types of gaze characteristics are related to the actions involved in the execution of optimal problem solving. This would also allow for the exploration of specific types of gaze characteristics that differ from those seen in optimal problem solving and how they lead to poor outcomes (i.e., non-optimal problem solving, time expired errors).

4.5 CONCLUSION

Our findings suggest that the use of traditional gaze behaviour analysis in conjunction with gaze-based entropy measures provides important temporal and spatial information related to the underlying cognitive processes supporting high-level visuospatial problem solving. Characteristic gaze biases within the workspace and goal space during optimal performance reflect the increasing demands of task complexity on planning and visuospatial working memory. Data revealed that gaze patterns during the planning stage of non-optimal responses may be indicative of erroneous information extraction from the visuospatial arrays and impaired maintenance of solution sequences in visuospatial working memory. To correct for early errors in the planning stage, gaze behaviour in the execution stage of non-optimal trials is directed to extract and process information from the goal space; consequently, movement execution time increases to perform all other processes required to reverse erroneous movements and re-sequence problem solutions to achieve the task goal. By conducting an extensive examination of eye movements during problem solving, this study provides valuable insight into the use of detailed gaze metrics to objectively assess and disentangle the cognitive processes underlying

planning and problem solving in healthy adults. Furthermore, the gaze analyses applied in the present study provide a promising means to identify, correct, and optimize eye movement patterns to support improved performance on cognitive tasks relying on planning and problem solving.

Section 5 – *The effects of task difficulty on gaze behaviour during landing with visual flight rules in low-time pilots*

A version of this chapter has been published in the Journal of Eye Movement Research (Ayala et al., 2023) and has been reproduced here in accordance with the Creative Commons Attribution 4.0 International License.

Ayala, N., Zafar, A., Kearns, S., Irving, E., Cao, S., & Niechwiej-Szwedo, E. (2023). The effects of task difficulty on gaze behaviour during landing with visual flight rules in low-time pilots. *Journal of Eye Movement Research*, 16(1). <https://doi.org/10.16910/jemr.16.1.3>

5.1 INTRODUCTION

Modern aircraft cockpits present a complex human-machine interface where the success of the flight depends on the pilot's ability to select relevant information from multiple competing stimuli. The dense visual field of instruments and displays conveys information about the status of the aircraft in real-time and must be closely monitored. Notably, 70-80% of global aviation accidents are caused by human error (Shappell and Wiegmann, 2000) with a major contributing factor proposed to be ineffective pilot monitoring of the plane, especially during dynamic phases of flight (i.e., take-off, final approach, and landing) (Boeing, 2021; National Transportation Safety Board, 1994). It is crucial to understand a pilot's information processing abilities underlying successful performance. Eye-tracking provides a non-invasive method that reveals discrete cognitive processes and strategies used to facilitate behaviour (Atik and Arslan, 2019; Ayala et al., 2022; Cao et al., 2022; Hodgson et al., 2019; Martin et al., 2017; Shiferaw et al., 2019; Vickers and Williams, 2017). For example, eye movement measures have been used as an

important index for hazard perception in driving (Cao et al., 2022; Cvahte Ojsteršek and Topolšek, 2019; Ziv and Lidor, 2016), human-machine interaction and usability assessment (Jacob and Karn, 2003; Liu et al., 2021; Menekşe Dalveren and Cagiltay, 2018; Niu et al., 2020), the development of visual strategies in athletes (Vansteenkiste, et al., 2022), and pilot behaviour assessment in aviation (Peißl et al., 2018; Vlačić et al., 2020). As such, the current investigation sought to examine the utility of gaze behaviour metrics to objectively characterize information processing in low-time pilots (fewer than 300 flight hours) during a landing task, and how it is altered by task difficulty.

The pattern of fixations and eye movements used to sample visual information in our environment is collectively referred to as gaze behaviour (Kandel et al., 2012). Gaze behaviour is task-dependent and tightly linked to the underlying perceptual, cognitive, and motor processes associated with the selection and processing of relevant sensory information (de Brouwer et al., 2021). Therefore, it has a direct influence on the planning and execution of subsequent actions (Gonzalez and Niechwiej-Szwedo, 2016; Land and Hayhoe, 2001). Studies have demonstrated that specific gaze metrics including total dwell time, fixation frequency, scan path length, average saccade amplitude and fixation duration are associated with performance in various laboratory paradigms (Ayala et al., 2022; Hodgson et al., 2019; Lemonnier et al., 2014; Martin et al., 2017; Zelinsky & Sheinberg, 1997). For instance, task-relevant areas tend to be fixated longer as task difficulty increases suggesting that fixation location and dwell time are a proxy for the allocation of attention (Ayala et al., 2022; Hodgson et al., 2000; León et al., 2019). Additionally, high performers show significant fixation biases toward task critical areas compared to low performers (Hodgson et al., 2000; León et al., 2019). Fixation frequency, duration and scan path length tend to increase as a function of task difficulty, while saccade

amplitudes decrease (Andrzejewska and Stolińska, 2016; Ayala et al., 2022; Hodgson et al., 2000; Kaller et al., 2009; Nitschke et al., 2012; Zelinsky & Sheinberg, 1997) due to the increased scanning and processing required to problem solve and elaborate on longer solution sequences. Collectively, research to date supports that gaze measures provide objective insight into information processing that underlies differences in task difficulty and performance. However, a major limitation linked to these traditional gaze measures is that they are often time-averaging operations; thus, failing to make use of the information regarding the patterns and sequences of gaze behaviour. Such information may be critical to consider when evaluating gaze behaviour during complex occupational tasks and environments.

Gaze entropy represents one of the more established, advanced methods to quantifying the dynamic aspects of gaze behaviour (Shiferaw et al., 2019). Entropy (measured in bits of information) is defined as the average information or uncertainty associated with choice (Shannon, 1984). In the context of gaze behaviour, the complexity of an individual's gaze pattern is governed by the number of regions (i.e., choices) that are fixated (Batty et al., 2014). These regions are characterized by defining the relevant areas of interest (i.e., AOIs) in an environment. There are two measures of gaze entropy that take into consideration the location and sequence of those fixations to compute gaze complexity, namely stationary gaze entropy (SGE) and gaze transition entropy (GTE). SGE is a measure of gaze dispersion that is computed over a given viewing period. The more equally distributed (i.e., wider gaze dispersion) fixations are across the environment (i.e., AOIs), the higher the entropy. Thus, a high SGE reflects an exploratory mode of visual attention whereas a low SGE reflects a focal mode of attention (Shiferaw et al., 2019). GTE examines fixation sequence complexity through the analysis of gaze transition matrices (Shiferaw et al., 2019). GTE assumes that fixation locations in a scan sequence are better

predicted from current and previous locations through a conditional probability (Weiss et al., 1989). High GTE indicates complex pattern of sequential scanning behaviour, which typically involves more frequent switching between more AOIs (Ayala et al., 2022; Shiferaw et al., 2019). In contrast, low GTE reflects a more predictable scanning sequence with fewer fixation transitions between fewer AOIs. Note that a low GTE value can signal two different scenarios, either gaze has become more efficient and directed to relevant AOIs or it may indicate a failure to properly monitor the task environment. To interpret the entropy findings appropriately, it is important to ensure that SGE and GTE are examined in conjunction with traditional gaze metrics and behavioural performance. Such an approach provides a comprehensive account of gaze behaviour in complex environments, and potentially more insight into the underlying neurocognitive and sensorimotor processes (Ayala et al., 2022).

Seminal research in the aeronautical domain has demonstrated an association between gaze behaviour and flying performance (for review see Glaholt, 2014; Peißl et al., 2018; Ziv, 2016). For instance, deploying attention to the external environment in visual flight rules (VFR) conditions provides the operator with relevant visual information that facilitates successful landing (Di Nocera et al., 2007; Gray et al., 2014; Kim et al., 2008; Sarter et al., 2007). Specifically, the optical splay angle is a visual cue that can be used to effectively align an aircraft with the runway centerline (Beall and Loomis, 1997), whereas the runway length-width ratio can be used to regulate altitude (Mertens and Lewis, 1981). These findings are made more apparent during night landings, when these cues are less perceptible and consequently impact performance (Kim et al., 2008). In addition to the use of visual cues, other gaze-specific parameters of attention allocation and information processing have been investigated through the use of traditional (i.e., dwell time, fixation frequency, fixation duration, saccade amplitude) and, more

recently, advanced gaze metrics to characterize pilot gaze behaviour and the extent to which visual scanning changed with cognitive load (Allsop and Gray, 2014; Allsop et al., 2017; Babu et al., 2019; Diaz-Piedra et al., 2019; Lounis et al., 2021; Tole et al., 1982), situational awareness (Dehais et al., 2017; van Dijk et al., 2011; van de Merwe et al., 2012) and level of expertise (Brams et al., 2018; Glaholt, 2014; Lounis et al., 2021; Peißl et al., 2018; Tole et al., 1982; Ziv, 2016). The advantage of eye-tracking to probe pilot characteristics is that it provides real-time, objective data with minimal interruptions to the experiment or user, unlike questionnaires and probes.

The goal of the current study was to expand on previous work that specifically examined changes in gaze behaviour as a function of task difficulty (Babu et al., 2019; Dick, 1980; Harris et al., 1986; Tole et al., 1982). A limitation in the previous literature is that cognitive load has been used interchangeably with workload, task load and task difficulty to describe the relationship between the demands imposed by a task and the availability of cognitive resources to perform that task. Here, we make a distinction between task difficulty and cognitive load. Task difficulty does not necessarily coincide with an increase in cognitive load, as the latter would be linked to task manipulations that increase the amount of information being held in working memory (Paas and van Merriënboer, 2020). Instead, task difficulty, which is germane to the current study, is tied to an increase in the sensorimotor control required to perform the task. For example, increasing the difficulty of a landing task by imposing high winds, shorter runways, and steep approaches due to terrain requires the pilot to impose higher sensorimotor control via corrective maneuvers during the operation of the plane in order to ensure a smooth, consistent, and safe flight (Federal Aviation Administration, 2021). Notably, there is a lack of consensus regarding the effects of task difficulty on gaze behaviour. While some studies showed

task difficulty had no significant impact on pilot scanning behaviour or performance (Dick, 1980; Dick, 1976; Krebs and Wingert, 1976), other work showed increased fixation frequency (Babu et al., 2019; Harris et al., 1986), more gaze transitions between task-relevant instruments (Waller, 1976), and longer dwell times on the runway (Di Nocera et al., 2007; Sarter et al., 2007). The conflicting findings reported in previous work are likely a consequence of the various methods employed to characterize scanning behaviour as well as the wide range of flying experience seen across the recruited participants (i.e., commercial pilots, military pilots, individuals with no flight experience).

The current study aimed to clarify previous findings by examining the effect of task difficulty in low-time pilot performance. As such, we systematically manipulated task difficulty during a simulated landing scenario and examined traditional and advanced gaze metrics. The current investigation differs from previous work in two important respects. First, the pilot group recruited here involves low-time pilots who are at the early stages of their training (i.e., ab initio pilots). Previous work that focused on gaze and task difficulty examined experienced military and commercial pilots in advanced aircraft configurations (i.e., large, multi-engine aircraft with glass cockpit displays) (Brams et al., 2018; Diaz-Piedra et al., 2019; Dick, 1980; Krebs and Wingert, 1976; van De Merwe et al., 2012; van Dijk et al., 2011; Vlačić et al., 2020). Gaze behaviour and sensorimotor control are significantly influenced by level of expertise (Abernethy, 1996; Burris et al., 2020; Gegenfurtner et al., 2011; Peißl et al., 2018; Ziv, 2016). Thus, it is important to investigate the relationship between task difficulty and gaze behaviour in pilots, especially during their initial stages of training when they have little to no flight experience (i.e., ab-initio pilots). Such knowledge will advance our understanding on the relationship between eye movements and task difficulty, with specific implications for developing pilot training

programs and evaluations. This is particularly relevant as improvements in the development of pilot competence in training are critical, since pilots are expected to progress more quickly from training through to airline and more advanced roles to address the expected international shortages of pilots (Kearns, 2021). Second, the current investigation examined the utility of using a comprehensive set of eye movement analyses to characterize gaze behaviour dynamics as a non-invasive means to probe how gaze and, by proxy, information processing are impacted by task difficulty.

In line with previous findings (Di Nocera et al., 2007; Harris et al., 1982; Sarter et al., 2007; Waller, 1976; Tole et al., 1982), we hypothesized that an increase in task difficulty (i.e., turbulent weather conditions) would be associated with an increase in dwell time (specifically outside the cockpit), higher fixation rate, and a reduction in SGE and GTE. These findings are expected to underlie a greater need to devote more time and attention toward-task relevant AOIs during turbulent conditions to extract critical information and ensure a safe landing.

5.2 MATERIALS AND METHODS

5.2.1 Participants

Eighteen participants were recruited from the student and alumni populations at the University of Waterloo (14 males, 4 females; age range: 18-25 years, mean=20 years old, SD=2 years). All participants were either current aviation students or had graduated from the aviation program (number of flight hours range: 0-280, mean=64 hours, SD=91 hours; PC flight simulator experience range: 5-100 hours, mean=37 hours, SD=31 hours). All participants had normal or corrected-to-normal vision and had not been previously diagnosed with a neuropsychiatric/neurological disorder or learning disability. Participation in the study was

voluntary, and participants received course credits as compensation. The study's protocol was approved by the University of Waterloo Research Ethics Board Committee (#43238), performed in accordance with the 2008 Declaration of Helsinki, and consent was obtained prior to beginning the protocol.

5.2.2 Apparatus

Participants sat in a height-adjustable chair with their chin placed in a chin rest. A 20-in LED monitor (85 Hz refresh rate 1920x1080 pixels, LG) was located at participants' midline with a viewing distance of 50 cm and was used to present visual stimuli (i.e., the flying scenarios). A second computer monitor (85 Hz refresh rate, 1024x768 pixels, View Sonic) that was only visible to the experimenter was used to record eye position data using the EyeLink II eye-tracker (SR Research Ltd, Ottawa, ON, Canada) sampling at 500 Hz. Participants used a joystick and throttle (TCA Officer Pack Airbus Edition, ThrustMaster, USA), placed beneath the simulation display, to provide all necessary input commands. Prior to data collection, a nine-point calibration of the eye tracker was performed. An immediate follow-up validation of calibration accuracy was conducted to verify that the error was $<1^\circ$ for each point in the calibration grid. Stimuli presentation and behavioural data acquisition were controlled using Microsoft Flight Simulator 2020 (Asobo Studio, France).

5.2.3 Scenario and task

Participants were tested in a single session (approx. 90 minutes). A visual screening was first completed including a visual acuity test using the Bailey-Lovie chart and a stereoacuity test using the Randot Stereo test (Stereo Optical Company, Inc.). Prior to commencing the experimental trials, training was performed to familiarize the participants with the joystick and throttle controls (TCA Officer Pack Airbus Edition, ThrustMaster, USA). The experimental

landing simulations were programmed in the Microsoft Flight Simulator landing challenge environment configured as a Cessna 152 (included steam-gauge instruments) flying into Billy Bishop airport (Toronto, ON, Canada). Participants were asked to complete a total of 20 customized landing challenges while their eye movements were recorded. The landing challenges were pseudo-randomized into 10 easy (i.e., high visibility and low wind conditions) and 10 difficult (i.e., high visibility and high wind conditions) trials. All participants received the exact same environmental configurations. **Figure 5.1** shows a screen capture of the simulated scenario. Each trial was pre-set to start as a straight-on approach to the airport at an altitude of 1000 ft, 2.5 nautical miles away from the runway with flaps and trim set to zero, and at a starting speed of 120kts. The simulated landing task involved visual flight rules (VFR) where visibility is high and represents one of the most basic landing scenarios that ab initio pilots are faced with during training. This allowed for the extension of previous work that used similar paradigms and more advanced aircraft configurations (i.e., helicopter simulators, A320 flight simulators, larger aircrafts with glass cockpit displays) (Brams et al., 2018; Diaz-Piedra et al., 2019; van De Merwe et al., 2012; van Dijk et al., 2011). This was particularly important as the present work recruited low-time pilots.



Figure 5.1. Illustration of the visual stimuli employed in the Microsoft Flight Simulator landing challenge environment. The participants point of view of the cockpit replicated that of a pilot flying a Cessna 152, pre-set for a straight-on approach to Billy Bishop airport, Toronto, Canada. The orange boxes represent the ten main areas of interest used in the gaze analyses. These include the airspeed (1), attitude (2), altimeter (3), turn coordinator (4), heading (5), vertical speed (6) and power (7) indicators, as well as the runway (8), horizon (9), and side window (10).

At the start of each trial, participants were asked to look at a red dot marked on the monitor in order to standardize initial eye position. The trial was then manually initiated after a drift correction was completed by the examiner. The goal of the task was to land the plane as smoothly and accurately as possible relative to a blue landing ‘goal’ box near the start of the

runway. The trial was terminated after the participant brought the plane to a complete stop, or if the landing was deemed unsuccessful (i.e., plane crash or plane landed off the runway).

5.2.4 Data reduction

Eye movement data were analysed offline using the eye tracker's Data Viewer software (ver 1.8: SR Research, Ontario, Canada). Eye-movement traces were visualized by the experimenter and played back at a slowed speed superimposed over the image displaying the gaming environment. The task environment was discretized within Data Viewer by organizing the gaming environment into ten areas of interest (AOIs) (**Figure 5.1**). The AOIs were manually defined to represent seven main gauges of interest within the cockpit including, airspeed (1), attitude (2), altimeter (3), roll coordinator (4), heading (5), vertical speed (6) and power (7). Three additional AOIs were also defined outside the cockpit including, the runway (8), the horizon (9), and the side window (10). Fixations found outside these AOIs were defined as a non-area of interest and excluded from the analysis (<3%). Based on previous work and discussions with higher-hour pilots (i.e., >800 hours) and instructor pilots, the AOIs defined in the current study were generated. In this way, the margin/border of each AOI is clearly explainable and allows for the entropy values to provide the most relevant and interpretable information from the scan patterns observed. This approach serves to be the most suitable for the current application because it is clear how the visual field should be grouped, and because there is strong ecological support from the piloting task and the cockpit design. In contrast, a grid/agnostic AOI approach would be most suitable for cases where researchers do not know how meaningful information in the visual field should be grouped. Eye movements were detected using a saccade detection algorithm implemented in Data Viewer with a 30°/s velocity

threshold and an $8000^\circ/\text{s}^2$ acceleration threshold. Fixations were defined as pauses between saccades that had a minimum duration of 80 ms (Krejtz et al., 2014; Velichkovsky et al., 2005). The current study focused on primary saccades, thus microsaccades ($<1^\circ$) were excluded from analysis (Martinez-Conde et al., 2013). Trial with missing data (i.e., loss of signal $>30\%$) ($\sim 5\%$ of trials) and outliers for each of the dependent variables (i.e., >1.5 the interquartile range around the first and third quartiles) ($\sim 16\%$ of trials) were removed.

5.2.5 Entropy analysis

The entropy-based analysis was completed using the ten AOIs (**Figure 5.1**) that were discretized during pre-processing. Eye fixations in the ten AOIs were assigned a number from 1 to 10 indicating the AOI where the eyes fixated. A sequence of fixation locations was then generated for each trial. Custom scripts were written in Python to compute both SGE and GTE.

SGE was computed by first producing a vector, $\mathbf{V}$, of length 10, where $\mathbf{V}_i$ was the total number of fixations at AOI i . $\mathbf{V}$ was then divided by the total number of fixations in the sequence, so that $\mathbf{V}_i$ was the probability of a fixation landing at AOI i . The probability vector $\mathbf{V}$ was then applied to **Equation 1** (Shannon, 1984).

$$H_{SGE}(V) = - \sum_{v \in V} v \cdot \log(v)$$

Equation 1

GTE was computed by first creating a 10x10 transition matrix, $\mathbf{M}$, where $\mathbf{M}_{i,j}$ was the total number of transitions from AOI i to AOI j . Each row, $\mathbf{M}_{i,*}$, was divided by the sum of row i , so that $\mathbf{M}_{i,*}$ represented the probability of fixation transition from AOI i to any of the ten AOIs. Finally, GTE was computed using **Equation 2** (Ciuperca and Girardin, 2007), applying the transition matrix $\mathbf{M}$ and the probability vector $\mathbf{V}$.

$$H_{GTE}(M) = - \sum_{i=1}^{10} V_i \sum_{j=1}^{10} M_{i,j} \cdot \log(M_{i,j})$$

Equation 2

5.3 RESULTS

Participants' behavioural performance and eye movements were analyzed while they were performing the landing task. All performance measures were provided by the gaming software at the end of every trial. Raw eye-movement data were provided by EyeLink software for the duration of each trial. The main hypothesis was tested using a one-way repeated measures ANOVA with task difficulty (easy, difficult) as the independent variable. Our analysis is divided into three parts. In the first part, the landing performance measures were examined as a function of task difficulty. In the second part, traditional gaze metrics were assessed as a function of task difficulty. In the third part, entropy-based gaze analyses were carried out as a function of task difficulty. All ANOVAs were performed with an alpha level set at 0.05. The Bonferroni post hoc correction for multiple comparisons was applied for all post hoc analyses to determine significant differences between variables.

5.3.1 Performance measures

Performance-dependent variables included success rate (%), completion time (sec), overall performance score (maximum of 2,000,000), landing accuracy (ft), ground roll (ft), and landing smoothness (feet per minute: fpm). These were all derived from Microsoft Flight Simulator 2020 (Asobo Studio, France). Success rate (%) was defined as the percentage of successful landing trials (i.e., participants landed on the runway without crashing) out of the total number of landing trials. Unsuccessful trials were automatically detected by Microsoft Flight

Simulator when an aircraft either crashed or landed off the runway. Notably, these trials were not analyzed further for performance or gaze measures due to their rare occurrence (<2% trials). All other parameters reported are based on successful landing trials. Completion time (sec) was defined as the duration from landing challenge onset to landing challenge offset. Landing challenge offset was automatically determined based on when the plane came to a complete stop on the runway. Overall performance scores were generated by Microsoft Flight Simulator software for every landing using **Equation 3**, which was dependent on three sub-scores (i.e., landing smoothness, landing accuracy, and ground roll). Landing smoothness (fpm) was defined as the sink rate just before and at plane touchdown. Landing accuracy (ft) was defined as the distance between the centerline of the runway and the plane's touchdown. Ground roll (ft) was defined as the distance between the center of the touchdown zone to the plane's full stop.

Overall Performance

$$= (\text{Accuracy score} + \text{Ground Roll score}) * \text{Landing Smoothness score}$$

Equation 3

As expected, landing success rate was lower for difficult trials (mean= 95%, SD= 8%) compared to easy trials (mean=100%, SD=0%) (**Figure 5.2A**). Completion time (sec) produced a main effect of task difficulty, $F(1,17)=105.740$, $p<0.0001$, $\eta_p^2=0.861$. **Figure 5.2B** demonstrates how difficult trials (mean=146 sec, SD=11 sec) took significantly longer to complete than easy trials (mean=132 sec, SD=9 sec). Moreover, overall performance yielded a main effect of task difficulty, $F(1,17)=113.456$, $p<0.0001$, $\eta_p^2=0.870$. Overall performance scores were lower during difficult trials (mean=903898, SD=299342) compared to easy trials (mean=1280915, SD=284673) (**Figure 5.2C**).

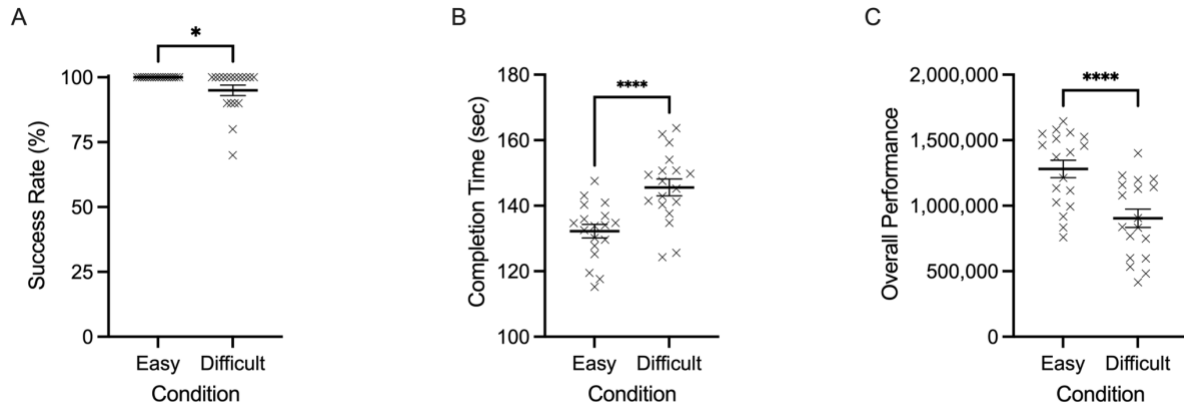


Figure 5.2. Individual data points and their respective group means for success rate (%) (A), completion time (sec) (B), and overall performance scores (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Landing smoothness (fpm) did not yield any significant effects $F(1,17)=2.371$, $p=0.142$ (**Figure 5.3A**). In contrast, landing accuracy (ft) yielded a main effect of task difficulty, $F(1,17)=22.024$, $p<0.001$, $\eta_p^2=0.564$. Results indicate that participants landed the plane with more lateral error (i.e., off-center) during difficult trials (mean=73 ft, SD=60 ft) compared to easy trials (mean=23 ft, SD=24 ft) (**Figure 5.3B**). Last, ground roll (ft) was not significantly modulated by task difficulty, $F(1,17)=0.176$, $p=0.680$ (**Figure 5.3C**).

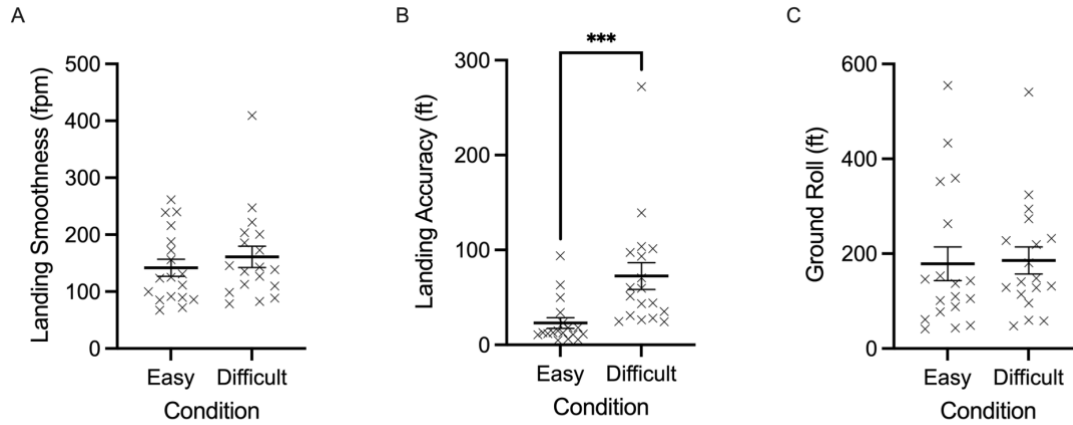


Figure 5.3. Individual data points and their respective group means for landing smoothness (fpm) (A), landing accuracy (ft) (B), and ground roll (ft) (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

5.3.2 Traditional gaze measures

Traditional gaze-based analysis was completed using the ten AOIs (**Figure 5.1**) that were discretized during pre-processing using Data Viewer (SR Research, Ontario, Canada). Gaze-dependent variables for the ten AOIs (**Figure 5.1**) included: dwell time (%), fixation rate (fixations/sec) and fixation duration (ms). Total scan path length ($^{\circ}$), average saccade amplitude ($^{\circ}$) and saccade amplitude ($^{\circ}$) variability were calculated across all AOIs. Dwell time (%) was defined as the total duration spent within a given AOI, which was converted to a percentage (i.e., with respect to total time). Fixation rate (fixations/sec) was defined as the number of fixations that occurred relative to the total time spent completing the landing challenge. Fixation duration (ms) was defined as the average duration of all fixations within a given AOI. Scan path length ($^{\circ}$) was defined as the sum of all saccade amplitudes. Last, average saccade amplitude ($^{\circ}$) was the

mean of all saccade amplitudes recorded, whereas saccade amplitude (°) variability was the within-participant standard deviation of saccade amplitude (°). Means and standard deviations for all traditional gaze measures are reported in **Table 5.1**.

Table 5.1. Traditional gaze values calculated for all areas of interest during easy and difficult conditions.

Task Difficulty	Dwell Time (%)		Fixation rate (fixations/sec)		Fixation duration (msec)	
	Easy	Difficult	Easy	Difficult	Easy	Difficult
AIRSPPEED	15.8(6.8)**	13.3(5.9)	0.58(0.24)**	0.48(0.20)	276(42)	280(54)
ATTITUDE	2.2(1.9)	1.8(1.6)	0.17(0.32)	0.08(0.06)	201(37)	200(42)
ALTIMETER	2.9(2.4)	2.6(2.5)	0.26(0.58)	0.11(0.09)	228(41)	230(62)
ROLL						
COORDINATOR	1.1(0.8)	0.93(0.7)	0.05(0.04)	0.04(0.03)	251(83)	221(58)
HEADING	0.8(0.5)	0.6(0.4)	0.23(0.71)	0.03(0.02)	210(74)	196(60)
VERTICAL						
SPEED	1.9(1.4)	1.8(1.3)	0.08(0.06)	0.07(0.05)	246(60)	249(68)
POWER	6.0(3.5)**	4.5(3.3)	1.16(3.58)	0.16(0.11)	264(51)	263(40)
	68.0(9.7)***					
RUNWAY	*	73.9(9.0)	6.27(18.21)	2.10(0.49)	357(47)	368(48)
HORIZON	3.8(2.1)**	2.7(1.8)	0.15(0.08)***	0.09(0.06)	254(47)**	295(73)
SIDE WINDOW	0.8(0.4)	0.9(0.5)	0.04(0.02)	0.03(0.02)	236(79)	249(101)

Note. Mean (standard deviation) values for all traditional gaze measures across all areas of interest (AOI) and task conditions (easy versus difficult). Significant changes between task difficulties and their corresponding dependent variable and AOI reported via $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$, $****p \leq 0.0001$.

Dwell time (%) revealed significant changes across several AOIs as a result of task condition. First, decreased dwell time was found for airspeed $F(1,17) = 13.006$, $p = 0.002$, $\eta_p^2 = 0.433$; power $F(1,17) = 13.043$, $p = 0.003$, $\eta_p^2 = 0.532$; and horizon AOIs $F(1,17) = 13.702$, $p = 0.002$, $\eta_p^2 = 0.446$. Runway dwell time (%) was the only AOI to show a significant increase as a function of task difficulty, $F(1,17) = 37.559$, $p < 0.0001$, $\eta_p^2 = 0.688$. All other AOI's did not reveal a significant change in dwell time, $F_s(1, 17) < 2.861$, $p_s > 0.117$ (**Table 5.1**). Note that the significant changes in the distribution of attention (i.e., dwell time %) observed between easy and difficult conditions are illustrated in **Figure 5.4**.



Figure 5.4. Illustration of the visual stimuli employed during landing and the corresponding change in dwell time (%) between easy and difficult conditions. Only significant changes in dwell time are displayed over their corresponding area of interest (AOI) (i.e., airspeed, power, runway, and horizon AOIs). Group mean differences (difficult-easy) are illustrated in red to indicate a decrease in AOI dwell time and green to indicate an increase in AOI dwell time.

Airspeed fixation rate demonstrated a main effect for task difficulty, $F(1,17)= 13.688$, $p=0.002$, $\eta_p^2=0.446$. Fixation rate within the airspeed AOI was higher during easy trials compared to difficult trials. Horizon fixation rate was also significantly modulated by task difficulty, $F(1,17)=22.335$, $p<0.001$, $\eta_p^2=0.568$. Easy trials had higher fixation rates within the horizon AOI compared to difficult trials. All other AOI's were not significantly modulated by task difficulty, $F_s(1,17)<1.436$, $ps>0.247$ (**Table 5.1**).

Horizon was the only AOI where fixation duration was significantly modulated by task difficulty, $F(1,17)=8.771$, $p=0.009$, $\eta_p^2=0.340$. Fixation durations were longer during difficult trials compared to the easy trials. All other AOIs did not have significantly altered fixation durations, $F_s(1,17)<2.177$, $p_s>0.056$. Lastly, scan path length was not significantly influenced by task difficulty, $F(1,17)=0.001$, $p=0.974$ (easy: 1205 °, SD=715°; difficult: 1200°, SD=546°). However, increased task difficulty was associated with a significant reduction in average saccade amplitude (easy: 4.96°,SD= 1.38°; difficult: 4.58°, SD=1.35°; $F(1,17)=10.078$, $p=0.006$, $\eta_p^2=0.372$) and saccade amplitude variability (easy: 6.53°,SD= 1.50°; difficult: 5.91°, SD=1.62°; $F(1,17)=16.951$, $p=0.001$, $\eta_p^2=0.499$).

5.3.3 Entropy-based measures

SGE represents the dispersion of fixations and an overall uncertainty of fixating in a particular AOI at any given moment (**Equation 1**). A higher SGE value represents a more spatially dispersed distribution of fixations across the AOIs. GTE represents the overall uncertainty associated with the temporal sequence of fixations, given the current fixation location (i.e., AOI), (**Equation 2**). Specifically, a higher GTE value indicates that gaze scan paths are more complex and frequently cross various AOIs in varying order throughout task completion.

SGE revealed a main effect of task difficulty, $F(1,17)=20.898$, $p<0.001$, $\eta_p^2=0.551$. Specifically, the spatial distribution of fixations was significantly more dispersed during easy trials (mean=1.7 bits, SD=0.4 bits) compared to difficult trials (mean=1.5 bits, SD=0.4 bits) (**Figure 5.5A**). Additionally, GTE produced a main effect of task difficulty, $F(1,17)=23.986$, $p<0.001$, $\eta_p^2=0.585$. **Figure 5.5B** shows how the overall gaze sequence was more complex due to an increase in AOIs being fixated in a more random sequence during easy trials (mean=1.3

bits, SD=0.3 bits) and became more predictable during difficult trials (mean=1.1 bits, SD=0.3 bits).

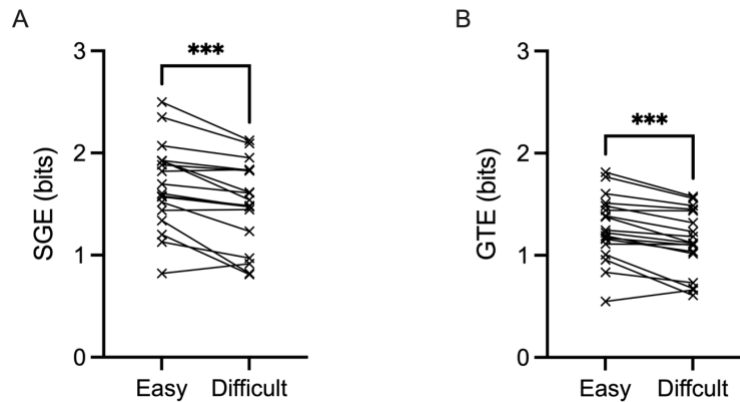


Figure 5.5. Individual data points for stationary gaze entropy (SGE) (bits) (A) and gaze transition entropy (GTE) (B) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Video recordings of participants' eye movements demonstrated that some participants adopted a gaze pattern that reflects a continuous allocation of attention outside of the cockpit, toward the runway. In contrast, other participants showed a gaze pattern that continually cycled through the various AOIs inside and outside the cockpit. Similar observations have been reported in other work (Allsop and Gray, 2014; Xiong et al., 2016); however, they were not objectively quantified. Therefore, additional analyses were completed to further examine how gaze entropy changed over the course of a trial. A 30-second average sliding window was initially chosen arbitrarily (Shiferaw et al., 2018); however, we found it did not have sufficient resolution to characterize the dynamic cyclical nature of the gaze patterns that were apparent in video

recordings of participants' gaze. As such, we used a 10-second average sliding window to produce a time trace that reflected how SGE and GTE evolved over time. To further characterize the observed pattern of switching gaze between the internal cockpit environment and external scenery, all cockpit AOIs (i.e., 1-7) were collapsed into a single 'inside' AOI and all external AOIs (8-10) were collapsed into as a single 'outside' AOI. The resulting entropy traces provide a temporal window depicting gaze dispersion and sequence predictability changes as participants completed the landing task (**Figure 5.6AB**).

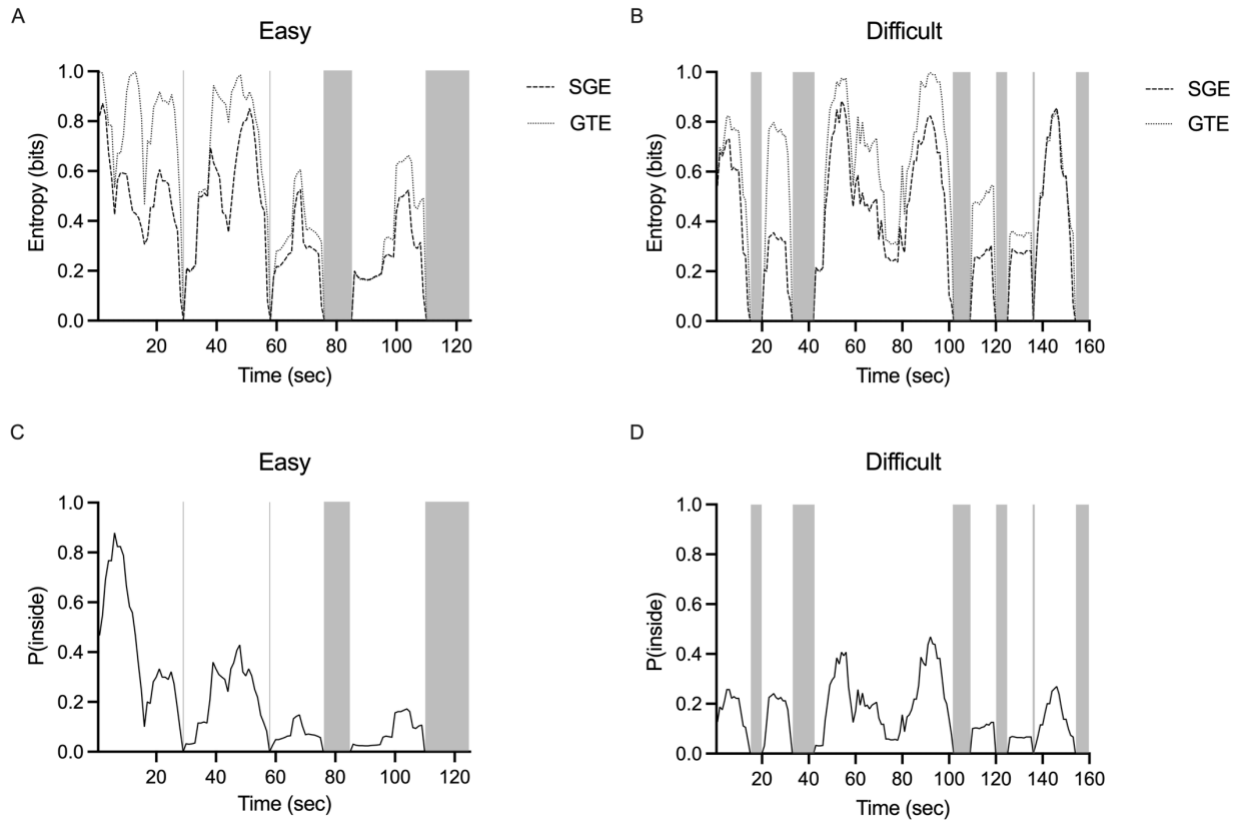


Figure 5.6. Representative time traces that reflect the stationary gaze entropy (SGE) and gaze transition entropy (GTE) of a single easy (A) and difficult (B) trial. Exemplar time traces showing the probability of fixating inside the cockpit for the easy (C) and difficult (D) conditions. These reflect the same easy (A and C) and difficult (B and D) trials. Gray boxes represent the sections of the trial where entropy and $P(\text{inside})$ equals zero. This indicates that gaze is directed outside the cockpit for at least 10 seconds (i.e., a single bout). The easy condition illustrates four bouts with a total bout time of 68 sec (A). The difficult condition illustrates six bouts with a total bout time of 96 sec (B).

Examining the SGE and GTE time traces revealed moments where both entropy measures fell to zero, which indicates that gaze was directed either outside or inside the cockpit for at least 10 seconds (**Figure 5.6AB**). To determine if the reduction in gaze entropy to zero reflected gaze allocation inside or outside the cockpit, we calculated the probability of a fixation being inside the cockpit over the length of the trial, $P(\text{inside})$ (**Figure 5.6CD**). Specifically, each fixation was assigned a binary number based on whether the fixation was inside or outside the cockpit. $P(\text{inside})$ was then computed as the number of fixations inside the cockpit divided by the total number of fixations in the 10 second window. When $P(\text{inside})$ was equal to 1, the participant was continuously fixating inside the cockpit for at least 10 seconds. When $P(\text{inside})$ was equal to 0, the participant was continuously fixating on the outside scenery for at least 10 seconds (**Figure 5.6CD**). This analysis provides an objective approach to monitor the temporal dynamics of gaze behaviour, which may reflect attention and how it is deployed inside and outside the cockpit. More specifically, the ‘bout’ analysis served to pinpoint the times when pilots stopped cycling through both internal and external environments. Given the time series of fixation probabilities showed distinct periods of fixations outside the cockpit, a ‘bout’ was defined as a period of time in which fixations remained entirely outside of the cockpit for at least 10 seconds. These bouts were detected as connected components (subsequent values) of zeros in the probability time series. Number of bouts was defined as the number of instances a bout was detected within a trial. Bout duration (sec) was defined as the average duration of all bouts that occurred in a trial. Total bout time (sec) was defined as the sum of all the individual bout durations within a trial.

Number of bouts yielded a main effect of task condition, $F(1,17)=7.224$, $p=0.016$, $\eta_p^2=0.298$. Specifically, easy trials (mean=2.0, SD=1) were associated with fewer bouts than

difficult trials (mean=2.4, SD=1) (**Figure 5.7A**). Average bout duration (sec) was not significantly impacted by task difficulty, $F(1,17)=4.233$, $p=0.055$ (**Figure 5.7B**). Last, total bout time (sec) was also significantly influenced by task difficulty, $F(1,17)=20.317$, $p<0.001$, $\eta_p^2=0.544$ (**Figure 5.7C**). Difficult trials (mean=45 sec, SD=26 sec) were associated with more continuous fixation outside of the cockpit compared to easy trials (mean=35 sec, SD=21 sec).

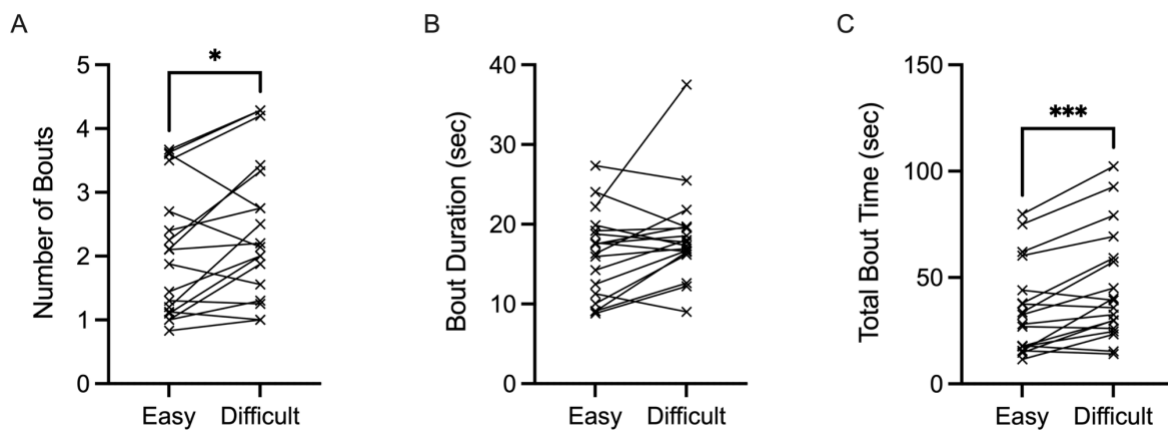


Figure 5.7. Individual data points for number of bouts (A), bout duration (sec) (B), and total bout time (sec) (C) are demonstrated for easy and difficult conditions. * $p\leq0.05$, ** $p\leq0.01$, *** $p\leq0.001$, **** $p\leq0.0001$.

5.4 DISCUSSION

This study characterized the effects of task difficulty on gaze behaviour and flight performance in low-time pilots as they completed simulated landing scenarios. Participants were asked to perform the landing task in high visibility, visual flight rules (VFR) conditions that differed based on the presence (difficult) or absence (easy) of strong winds. Several notable contributions emerged from this study. First, in support of our hypothesis, a comprehensive

assessment of gaze behaviour during landing revealed that dwell time and entropy-based measures were modulated by changes in task difficulty. Specifically, increasing task difficulty resulted in longer fixation of the runway, and a reduction in the dispersion (SGE) and complexity (GTE) of gaze sequences. Second, further exploration of the data led to a development of a novel approach to objectively identify and track instances when pilots selectively allocate their attention outside the cockpit.

5.4.1 The effect of task difficulty on performance

The analysis of landing performance provided unambiguous evidence about the successful manipulation of task complexity. Specifically, when dealing with strong wind conditions, pilot performance decreased. A finding that was made evident through a reduction in landing success rate and overall performance score, alongside an increase in completion time and landing accuracy error. Although a formal examination of subjective perception of task difficulty was not conducted in the current study, participants reported that they noticed an increase in task difficulty during the trials that involved turbulent weather conditions. Overall, these results are in line with earlier studies using similar experimental procedures (Diaz-Piedra et al., 2019).

5.4.2 The effect of task difficulty on pilot's gaze behaviour

The examination of gaze behaviour via traditional gaze metrics provides a proxy for the allocation of attention toward task-relevant information and how it is used to facilitate task performance (Di Nocera et al., 2007; Gray et al., 2014; Kim et al., 2008; Sarter et al., 2007; Ziv, 2016). The current study demonstrated that task difficulty was associated with a significant increase in the time spent looking outside toward the runway and a corresponding reduction in the time spent fixating on alternative AOI's that included the airspeed indicator, the power

indicator, and the horizon. These findings suggest that attention was focused on the runway during strong wind conditions at the expense of other AOIs both within and outside of the cockpit. In line with these findings, average saccade amplitude and variability also decreased as a function of task difficulty. A finding that indicates a reduction in the spread of sequential fixation locations, and hence a focusing of visual attention to a limited set of closely spaced regions within the task environment. Notably, the addition of strong wind conditions introduced a significant crosswind component which required continuous monitoring of heading direction during final approach and landing. The use of the crab or sideslip methods to compensate for the crosswind may have led participants to look outside more, not only to check on wind conditions but to monitor the plane's trajectory and ensure that the glideslope and heading were maintained and adjusted accordingly (Federal Aviation Administration, 2021). This is further supported by other work demonstrating that the allocation of attention outside the cockpit in VFR conditions helps provide the pilot with relevant visual information that facilitates successful landing (Beall and Loomis, 1997; Di Nocera et al., 2007; Gray et al., 2014; Kim et al., 2008; Mertens and Lewis, 1981; Sarter et al., 2007). Though fixation rates did not increase as a function of task difficulty across any AOIs, they did decrease for the airspeed indicator and the horizon. Further demonstrating that the scanning of information in these AOI's was reduced at the cost of spending more time fixating on or around the runway, which contributed to the reduction in dwell time percentages in these respective AOIs. These findings echo previous work by Van de Merwe et al. (2012), which found that fixation rates on various instruments were related to their problem-relevance; with the runway being of prime importance (Brown et al., 2002; Lu et al., 2020). Interestingly, a study by Babu et al. (2019) demonstrated that fighter pilots increased their fixation frequency when task difficulty increased during a longitudinal target tracking task, while

Tole et al. (1983) found no change in the fixation rate of military pilots associated with turbulence in a landing task. The contradictory findings may lie in the task employed (Babu et al., 2019) and the role that experience plays in the gaze behaviour supporting performance (Babu et al., 2019; Tole et al., 1983). For instance, experienced pilots tend to fixate relevant cockpit gauges for briefer periods of time more frequently compared to novice pilots (Glaholt, 2014). Additionally, our fixation rate does correct for the increased time spent completing the landing task during the turbulent conditions, whereas previous studies simply reported the fixation frequency count (Harris et al., 1986; Lu et al., 2020; Spady, 1978). Taken together, our traditional eye movement analysis measures suggest that the allocation of attention in low-time pilots became more biased toward the runway in order to monitor and extract the necessary information needed to land during challenging turbulent flight scenarios.

The analysis of gaze entropy is a surrogate for exploring the dynamic nature of gaze behaviour. Specifically, the dispersion (SGE) and sequence (GTE) of a scanning pattern have been shown to be effective indicators of attention deployment (i.e., focal versus exploratory modes of attention) as well as the complexity of the scanning pattern structure (Shiferaw et al., 2019). Indeed, previous work has shown that the competition of bottom-up (salience-driven) and top-down (executive control-driven) attentional processes influence the spatial prioritization of where we look (Bisley and Goldberg, 2010; Eysenck et al., 2007; Fecteau and Munoz, 2006; Shiferaw et al., 2019). For instance, bottom-up processes would likely drive attention to salient objects in the environment which may correspond to more random and dispersed fixation patterns (Shiferaw et al., 2019). Whereas top-down input would direct attention based on task knowledge, expectations, and current goals, and thus reflect task engagement via a reduction in the dispersion of fixations and more structured scanning patterns (Shiferaw et al., 2019). The

current findings revealed that low-time pilot's gaze entropy became less dispersed and less complex during strong wind conditions. That is, pilots followed a more predictable and deterministic visual scanning pattern during the more complex landing scenarios. This is analogous to previous studies that also reported an increase in the scanning and allocation of attention to a limited set of task-critical AOIs during more complex landing scenarios (Harris et al., 1986; Waller, 1976). Notably, these findings are similar to those seen using the saccade amplitudes metrics (Lemonnier et al., 2014). However, in addition to being able to capture the spread of fixations across task-relevant areas of interest that are object-defined (i.e., the runway, side window, specific cockpit gauges) the secondary entropy measure, GTE, helps to characterize the complexity of the scanning sequence itself within the task environment. This provides critical information regarding visual scanning patterns that can differ significantly despite occupying the same spatial regions. In line with previous work, our findings support the notion that pilots use exploratory and saliency-driven gaze patterns when the aircraft is in an error-free state (i.e., perfect flying conditions) and change their gaze behaviour when they are experiencing periods of flight that involve more challenging task demands; a shift that reflects the top-down attentional control imposed on visual scanning to help focus attention to the appropriate object at the appropriate time (Ayala et al., 2022; Bellenskes et al., 1997; Brams et al., 2018; Eysenck et al., 2007; Hogson et al., 2000; Shiferaw et al., 2019; Tole et al., 1983). It is also important to note that additional entropy analysis was conducted to ensure that AOI size disparities across the 10 defined regions did not bias the entropy results. We calculated entropy using the two AOIs (outside vs. inside), which resulted in a 50-50 screen split of equally sized AOIs. Indeed, the main effects reported in the current study were maintained and demonstrate that our entropy findings were not biased by AOI size. Specifically, SGE and GTE demonstrated

a main effect of task difficulty, $F(1,17)=6.652$ and 13.991 , $p= 0.019$ and 0.002 , $\eta_p^2= 0.281$ and 0.451

The current study makes an important contribution by quantifying the spatiotemporal gaze dynamics using a novel approach. Previous work reported an observation that novice pilots tend to restrict their gaze toward the external view of the cockpit and to focus predominantly on the runway, which has been referred to as ‘gaze tunneling’ (Xion et al., 2016; Ziv 2016).

Unfortunately, previous studies did not quantify this behaviour objectively. We adopted a sliding window approach to quantify the moment-to-moment changes in gaze entropy and the probability of fixating inside and outside the cockpit. This approach enabled the quantification and analysis of runway tunneling ‘bouts’ – that is, the continuous fixation of attention outside of the cockpit toward the runway, which were then examined with respect to their relationship with pilot performance and task difficulty. Our results demonstrated that an increase in task difficulty was associated with a greater number of bouts as well as an increase in total bout time.

Therefore, strong wind conditions resulted in low-time pilots becoming more stringent in attending to information outside of the cockpit- for intervals greater than 10 seconds- instead of inside toward the cockpit instrument panel; a finding taken to evince less cycling of attention between the external and internal cockpit environments. Notably, this may provide an alternative explanation for why entropy values decrease during more difficult trials. That is, instead of visual scanning becoming more structured and deterministic, it may be that SGE and GTE values are reduced due to an increase in gaze tunneling; an indicator of poor monitoring. In this scenario, it is expected that these gaze tunneling events will impair task performance. Though there were not enough trials to conduct a statistical analysis of these characteristic runway tunnelling bouts on failed landing trials, a preliminary analysis of these rare landing failures

demonstrated longer bout durations and total bout time during the early part of the flight in these trials. Additionally, we found that a group of less experienced low-time pilots ($n=11$) had more bouts of fixated attention outside the cockpit and longer total bout time compared to low-time pilots who had at least obtained their private pilot's licence (PPL) ($n=7$). Notably, the analysis of gaze behaviour profiles has not been examined in these smaller experience increments, and are being further investigated in a subsequent study. These initial findings suggest that the exclusive fixation of attention toward the runway for extended periods of time may impact pilots' situational awareness and mental models (Allsop and Gray, 2014; Federal Aviation Administration, 2021; Xiong et al., 2016). Though, neither of these outcomes would necessarily lead to poor landing performance when the plane is in an error-free state, this could pose a significant problem to landing performance and safety should an emergency/error arise. Further investigation of gaze behaviour is required as it could potentially lead to the development of an objective method allowing an 'online' detection of pilot inattention, poor pilot monitoring, or a general marker of divergence from optimal scanning.

5.4.3 Limitations

This study provides several notable contributions regarding how task difficulty may influence gaze behaviour. However, the current results are constrained by at least two limitations that should be addressed in future research. First, the methodological challenges imposed by the Microsoft Flight Simulator gaming environment limited real-time synchronization and made it difficult to understand the relationship between game generated landing performance scores and their ecological significance. Additional research using other simulator environments is needed to address this issue. For instance, it is crucial to address the basic relationship and temporal contingency between specific eye movements and actions (i.e., vision-in-action paradigms) to

determine whether specific types of gaze characteristics are related to superior performance of specific actions. Alternatively, this could also allow for the exploration of specific types of gaze characteristics that lead to poor outcomes (i.e., accidents, unstable approaches, loss of control). Second, we compared low-time pilot performance during an error-free state landing scenario in VFR conditions to that during strong wind conditions. It may be important to consider exploring the gaze profile differences between different classes of low-time pilots (i.e., those who have at least obtained their PPL and those who have not). Further research should take these smaller intervals of expertise into account starting from the ab initio stage (i.e., little to no flying experience) up to expert (i.e., every new licensing level- PPL, Commercial Pilot's License, etc.) with much larger sample sizes (i.e., >15 pilots per group). This will allow for a finer characterization and examination of gaze behaviour evolution and skill progression throughout the course of pilot training. This remains a crucial gap in the literature because little is known about the role that gaze behaviour plays in early skill learning and progression in complex environments.

5.5 CONCLUSION

This work highlighted the performance and gaze differences in low-time pilots when completing a simulating landing scenario in VFR conditions with and without turbulence. During turbulent weather conditions, pilots shifted their gaze behaviour to become less complex and more focused on the runway, thus making the scanning and processing of information more targeted towards fewer areas of interest. Overall, the gaze-related metrics used in the present study provide valuable information for assessing pilot gaze behaviour in the cockpit and can contribute to better characterization of visual scanning. This remains an important area of

research because understanding how gaze contributes to optimal pilot performance might be an important benchmark for monitoring and ensuring flight safety as well as evaluating pilot competency during training.

Section 6 – Investigating the role of flight phase and task difficulty on low-time pilot performance, gaze dynamics and subjective situation awareness during simulated flight

A version of this chapter has been published in the Journal of Eye Movement Research (Ayala et al., 2024) and has been reproduced here in accordance with the Creative Commons Attribution 4.0 International License.

Ayala, N., Kearns, S., Irving, E., Cao, S., & Niechwiej-Szwedo, E. (2024). Investigating the role of flight phase and task difficulty on low-time pilot performance, gaze dynamics and subjective situation awareness during simulated flight. *Journal of Eye Movement Research*.

6.1 INTRODUCTION

Most of the information regarding flight controls and operation is gathered and processed by the human visual system. Modern cockpits present a complex human-machine interface with multiple competing stimuli, thus pilots must be able to optimally scan and process critical pieces of information for the safe and successful operation of an aircraft. As such, eye-tracking research in the aeronautical domain has largely focused on the development and implementation of eye-tracking metrics with the objective of understanding what constitutes superior pilot monitoring of aircraft controls in expert pilots (Glaholt, 2014; Peißl et al., 2018; Ziv, 2016). Accordingly, several studies have already identified changes in traditional and advanced gaze (i.e., coordinated head and eye movements) metrics in expert pilots that vary with task difficulty, which are associated with differences in pilot performance and situation awareness (SA) (Brams et al., 2018; Glaholt, 2014; Robinski & Stein, 2013; for review see Peißl, 2018 and Ziv 2016). However, what remains is a specific gap in understanding how these measures are impacted within the confines of early-stage pilot skill development. Therefore, the aim of the current study

was to understand the impact of task difficulty on flight performance, gaze behaviour and SA across various phases of flight in low-time (i.e., <300 hours) pilots using a high-fidelity simulator environment.

Eye-tracking offers insight into skill performance by providing both behavioural and physiological outputs that can be used to provide insight regarding pilot performance, as well as the underlying cognitive processes (deBrouwer et al., 2021; Hermens et al., 2013; Land and Hayhoe, 2001; Robinski & Stein, 2013; Yarbus, 1967). For instance, it has been well established that high performing pilots spend more time selectively allocating their visual attention towards objects that are relevant to the task goals, while ignoring other task irrelevant areas (Ayala et al., 2023a; Di Nocera et al., 2007; Glaholt, 2014; Gray et al., 2014; Kim et al., 2008; Sarter et al., 2007; Van de Merwe, 2012). Specifically, optical splay angle and runway length-width ratio have been referenced as helpful runway visual cues that improve landing performance (Beall and Lumis, 1997; Kim et al., 2008; Mertens and Lewis, 1981). These apparent gaze biases are not only highly task dependent but demonstrate how high performing pilots efficiently and effectively scan their environment to sample all necessary information required to successfully plan and complete a specific task. What is important to note is that skill performance in itself may be very similar between any two pilots, or two groups of pilots. However, the way in which information is processed during a given task may differ despite their comparable performance capabilities. Eye tracking can help reveal these differences through the examination of other basic and dynamic gaze metrics that provide further insight into how efficiently individuals process information and could reveal the differing levels of task demands associated with a particular scenario. For example, previous work has shown that more efficient information processing is associated with reduced fixation duration (Andrzejewska & Stolińska, 2016; Brams

et al., 2018; Glaholt 2014; Peißl, 2018; Sun et al., 2016; Tang et al., 2016; Ziv, 2016), increased fixation frequencies (Brams et al., 2018; Gidlöf et al., 2013; Glaholt, 2014; Hebbar et al., 2023; Peißl, 2018; Ziv, 2016), as well as a greater propensity to fixate more task-relevant areas of interest (AOIs) - greater fixation dispersion (i.e., Stationary Gaze Entropy: SGE) - in a more variable/flexible pattern; as made apparent with greater fixation sequence complexity (i.e., Gaze Transition Entropy: GTE) (Ayala et al., 2022; Ayala et al., 2023a, 2023b; Glaholt, 2014; Hebbar et al., 2023; Peißl, 2018; Sun et al., 2016). Several studies have also demonstrated how reduced blink rate is a reliable proxy for increased task difficulty as well as cognitive load (Glaholt, 2014; Peißl et al., 2018). The application and exploration of these basic and dynamic gaze metrics remain important areas of inquiry with respect to pilot training and skill mastery because even if performance is optimal in an error-free state of flight, there may be serious ramifications to flight performance and safety should an emergency/error arise; especially, if such an event is associated with an overload in the pilots' information processing capabilities.

Studies examining gaze behaviour in pilots across various stages of flight demonstrated distinct gaze behaviour changes as task demands changed (Ayala et al., 2023b; Babu et al., 2019; Comstock, 1995; Dehais, 2020; Lijing et al., 2016; Lijing et al., 2014; Valcic et al., 2020). To our knowledge, only two studies have specifically investigated the impact of differing task demands on flight performance and gaze behaviour in low-time pilots (Ayala et al., 2023b; Dehais, 2020). Ayala and colleagues (2023) assessed gaze behaviour in 18 low-time pilots (flight hour range: 0-240, mean= 64 hours, SD= 91) as they completed simulated landing scenarios of varying difficulty (i.e., easy: no wind, high visibility; difficult: high winds, high visibility) programmed in a desktop computer Microsoft Flight Simulator game environment (2020, Asobo Studio, France). Results showed that an increase in task difficulty was associated with a longer

dwell time toward the runway, along with a reduction in fixation sequence dispersion and complexity. Moreover, prolonged fixation on a singular AOI (i.e., front window)- also known as cognitive tunneling (Bell et al., 2005; Engström et al., 2005; van Leeuwen et al., 2015)- became evident, demonstrating a reduction in pilot monitoring of internal cockpit gauges as a result of increased task difficulty. Therefore, it was concluded that gaze in low-time pilots became less complex and more focal; thus, making the scanning and processing of information more targeted toward task relevant AOIs when task difficulty increased. Similarly, Dehais and colleagues (2020) assessed gaze behaviour in 7 low-time pilots (flight hour range: 80-250 hours), as they completed two traffic patterns and basic flight maneuvers in a real aircraft. Results demonstrated significant changes in gaze dwell time patterns as a function of task demands (i.e., subgoals) imposed by different stages of flight (i.e., take-off, downwind, final approach). Although this study focused exclusively on characterizing the dwell time patterns across AOIs during different stages of flight, it was instrumental to showing how eye-tracking could be used for training purposes (i.e., informing pilot monitoring strategies) as well as highlighting the importance of accounting for stages of flight in gaze analytics.

The present study sought to expand on previous work in three important respects. First, the current study specifically focused on pilots with their Private Pilots' License (PPL) or Commercial Pilots' License (CPL). Earlier work demonstrated that there is a significant learning curve that occurs within the ab-initio stage of skill development that is associated with more variability around the performance and gaze measures of interest that may go unexplained and impact our conclusions (Ayala et al., 2023a). Furthermore, the use of high-fidelity flight simulators has been shown to immediately diminish the performance of ab-initio pilots who are unfamiliar with the immersive cockpit environment (Noble, 2002). As such, using strict

recruitment criteria to define the cohort was seen as a necessary step to reduce the heterogeneity across participants, and specifically understand how gaze, flight performance and SA interact as a function of task difficulty across flight phases in the cluster of pilots who are licensed to fly a single pilot aircraft. Second, the use of a high-fidelity flight simulator in the current study provides more data related to aircraft control and landing performance throughout the duration of the trial. Indeed, one of the limitations with using Microsoft Flight Simulator in previous work was that there were no continuous measures of aircraft control that could be examined in line with continuous measures of gaze behaviour. This limited the ability to examine the ongoing dynamic interactions between gaze and aircraft control during all phases of landing (i.e., downwind, base, final approach) in the simulated scenario. Lastly, the current investigation included an assessment of situation awareness in the form of a SA questionnaire (Situational Awareness Rating Technique: SART; Taylor, 1990) to assess the relationship between gaze behaviour and SA in low-time pilots. Although other questionnaires for SA exist (i.e., Situation Awareness Global Assessment Technique: SAGAT) (Endlsey, 1988), they require the pausing of the flying scenario to ask probing questions. This would significantly alter the gaze patterns being observed throughout the experiment and would prevent accurate characterization of gaze behaviours during each stage of flight.

In line with previous work (Ayala et al., 2023b), we hypothesized that in visual flight rules (VFR) conditions, an increase in task difficulty will result in increased dwell time and fixation counts toward the external environment (i.e., front window), increased cognitive tunneling (i.e., frequency and duration), and an overall reduction in the dispersion and complexity of visual scan patterns as indicated by entropy measures. It was expected these changes in gaze behaviour reflect a greater need to allocate more time and attention toward

fewer, more important task relevant AOIs during difficult conditions. It was also expected that the changes in gaze measures would be associated with reduced flight performance parameters and correlate with the pilot's level of experience (i.e., flight hours) and SA (Shiferaw et al., 2018; Svensson et al., 1997; Van de Merwe, et al., 2012).

6.2 METHODS

6.2.1 Participants

Twenty-four participants (male: 14; age range: 19-40 years, mean= 22 years old, SD= 4 years) were recruited from the student and alumni populations at the University of Waterloo. All participants were current aviation students or other individuals who had obtained at least their private pilot's license (PPL). All participants had normal or corrected-to-normal vision and had not been previously diagnosed with a neuropsychiatric/neurological disorder or learning disability. Participation in the study was voluntary, and participants received \$25/hour as remuneration. The study's protocol was approved by the University of Waterloo Research Ethics Board Committee (#43564), performed in accordance with the 2008 Declaration of Helsinki, and consent was obtained prior to beginning the protocol.

6.2.2 Experimental Setup and Apparatus

Flight Simulator. An AL250 ALSIM flight simulator (ALSIM, France) configured as a generic single engine aircraft that is representative of a Cessna 172 was used with the necessary instrument panel (steam gauge configuration), an avionics/GPS system, an audio/lights panel, a breaker panel, and a Flight Control Unit (FCU) (**see Figure 6.1**). Participants sat in a height-adjustable seat (left pilot seat) with their aviation headset plugged in for ATC (air traffic control)

callouts. The field of view covered by the simulator was 250° by 49° via panoramic VFR-VR-HD projectors. The participants controlled the aircraft with a yoke, throttle lever, and rudder pedals. Stimuli presentation, and behavioural data collection and acquisition were controlled from the Instructor Station and Engineering pack (ALSIM, France).

Eye Tracker. MindLink eye-tracking glasses (AdHawk Microsystems Inc., Waterloo, ON, Canada) were used to track the participants' eye and gaze movements (**Figure 6.1**). MindLink is a non-camera-based eye tracker embedded in a frame of eyeglasses that uses an ultra-compact micro-electromechanical system (MEMS) to track the eye and gaze movements (Zafar et al., 2023). The eye tracker was operating at 250 Hz, transmitting the gaze data and the video of its front-facing camera (82° field of view, 1080p, 30 Hz) via the AdHawk eye tracking software to a laptop (60 Hz refresh rate, 1920 x 1080 pixels, Microsoft 11) visible only to the experimenter (**Figure 6.1**).



Figure 6.1. Illustration of the simulator and eye tracking equipment set-up. The ALSIM simulator was set up as a single-engine aircraft, controlled with the yolk, throttle lever, and rudders pedals (A). Participants also used a headset (hanging on the left window) to make ATC calls (A). The AdHawk glasses (B) were worn by participants throughout the experimental session and were connected to the collection laptop (B) via a USB-C cord. The collection laptop used AdHawk software to calibrate the eye tracker and start and stop eye tracking data collection.

6.2.3 Scenario and task

Participants were tested in a single session (approx. 90 minutes). A visual screening was first completed including a visual acuity test using the Bailey-Lovie chart and a stereoacuity test using the Randot Stereo test (Stereo Optical Company, Inc.). Prior to commencing the experimental trials, a pilot briefing (completed by an instructor pilot) and practice trial was performed to familiarize the participants with the simulator environment and flight path (AL 250, ALSIM, France). The briefing also covered the segments of the flight that included the downwind (antiparallel), base (perpendicular) and final approach (in line with runway) phases of

flight, which are primarily determined based on the spatial orientation of the aircraft relative to the designated runway and the required tasks associated with each leg of flight (Federal Aviation Administration, 2021) (**Figure 6.2**).

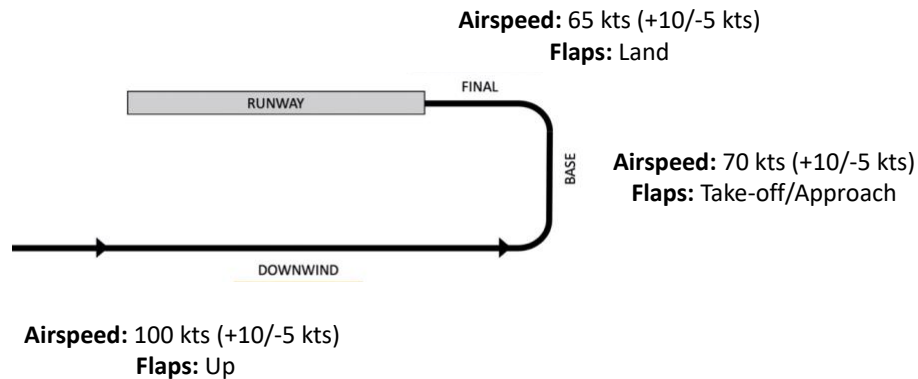


Figure 6.2. Illustration of the flight path and its respective phases of flight encountered in the experimental landing scenarios: Downwind, Base, Final Approach.

The experimental landing scenarios were programmed in the flight simulator environment, flying into the Region of Waterloo International Airport (CYKF; Runway 26), Breslau, Ontario, Canada. Participants were asked to complete a total of 8 customized landing trials while their eye and head movements were recorded. The landing challenges were pseudo-randomized into 4 easy (i.e., high visibility [>20 miles] and low wind [0 kts, 0°] conditions), and 4 difficult (i.e., high visibility [>20 miles] and high wind [26 kts, 230°] conditions) trials (i.e., 2 task difficulties/scenarios with 4 attempts/trials in each scenario). Note that the wind speed and direction for the difficult trial produced a cross wind component of 13 kts, while providing a tail wind of 22.5 kts during the downwind stage of flight and a head wind of 22.5 kts during the final approach stage of flight. This inherently allowed the downwind leg of flight to be completed

faster, while the final approach stage of flight was likely to be completed much slower, compared to the easy condition. All participants received identical environmental configurations. Figure 3 illustrates the visual conditions of the simulated scenario. Each trial was pre-set to start as a downwind-to-base-to-final approach to the airport at an altitude of 2017 ft at sea level (airport altitude 1054 ft), 1 nautical mile away from the downwind runway threshold with flaps and trim set to zero, and at a starting speed and power of approximately 110 kts and 2000 rpm, respectively. The simulated landing task involved VFR conditions where visibility was high, which represents one of the most basic landing scenarios that novice pilots are faced with during training. This allowed for the extension of previous work that used similar paradigms and more advanced aircraft configurations (i.e., helicopter simulators, A320 flight simulators, larger aircrafts with glass cockpit displays) (Ayala et al., 2023b; Brams et al., 2018; Diaz-Piedra et al., 2019; van De Merwe et al., 2012; van Dijk et al., 2011). This was particularly important as the present work recruited low-time pilots (i.e., <300 hours of flight time).

At the start of each trial, the participant manually initiated the landing scenario once 9-point eye-tracking calibration and validation procedures were completed by the examiner (average gaze error $<2^\circ$). The goal of the task was to land the plane as smoothly and accurately as possible relative to the center of the 500 ft markers near the start of the runway. The trial was terminated after the participant brought the plane to a complete stop, or if the landing was deemed unsuccessful (i.e., plane crash or plane landed off the runway). The participant was then asked to complete the Situational Awareness Rating Technique (SART) questionnaire to gauge their subjective opinion on various domains related to task difficulty, as well as the supply and demand of attentional resources required during task performance (Taylor, 1990).

6.2.4 Data reduction

Gaze data were post-processed offline using a custom-made script that used the 3D gaze vectors provided by the AdHawk software for saccade and fixation detection. The saccade detection algorithm was based on the algorithm proposed originally by Nyström and Holmqvist (2010) with some modifications to work on the current data captured at 250 Hz. Unlike the original method that uses a fixed saccade-peak-velocity threshold for detecting the saccade candidates, we used a low-pass filtered version of the velocity signal (rolling average with a 50-sample window) to increase the threshold in the noisy regions of the data. After classifying the eye movements into saccades and fixations, the average of the gaze sample during each fixation was taken as the fixation position. Eye-movement traces were visualized by the experimenter and played back at a slowed speed superimposed over the video displaying the simulator environment. The task environment was discretized using a custom code by organizing the simulator environment into ten areas of interest (AOIs) (**Figure 6.3**). The AOIs were manually defined to represent seven main gauges of interest within the cockpit including, airspeed (1), attitude (2), altimeter (3), VHF Omni Range (VOR) (4), heading (5), vertical speed (6) and power (7). Three additional AOIs were also defined outside the cockpit including, the front window (8), the left window (9), and the right window (10). Fixations found outside these AOIs were defined as a non-area of interest and excluded from the analysis (<4%). The current study focused on primary saccades, thus microsaccades (<1°) were excluded from analysis (Martinez-Conde et al., 2013). Trials with missing data (i.e., loss of signal >30%) (~4% of trials) and outliers for each of the dependent variables (i.e., >1.5 the interquartile range around the first and third quartiles) (~2% of trials) were removed.



Figure 6.3. Illustration of the visual stimuli employed in the AL250 flight simulator environment. The participants point of view of the cockpit replicated that of a pilot flying a Cessna 172, pre-set for a downwind-to-base-to-final approach to Waterloo International Airport, Breslau, Ontario, CA. The orange boxes represent the ten main areas of interest used in the gaze analyses. These include the airspeed (1), attitude (2), altimeter (3), VOR (4), heading (5), vertical speed (6) and power (7) indicators, as well as the front (8), left (9), and right windows (10).

6.2.5 Performance, Gaze, and Situation Awareness Analysis

We evaluated the flight performance, gaze behaviour, and subjective level of situation awareness across the two task difficulties (easy, difficult). Flight was assessed across two domains: 1) Landing Performance and 2) Aircraft Control. Landing Performance included

completion time (sec; duration of time from the start of the landing scenario to the plane coming to a complete stop on the runway), landing accuracy (degrees; the difference between the center of the plane and the center of the 500 ft runway marker at point of touchdown), and landing hardness (feet per minute, or fpm; the rate of decent at point of touchdown). Aircraft Control included the mean, standard deviation (i.e., variability) and root mean square error (RMSE) of the aircraft airspeed (i.e., the average difference between the reference optimal airspeed [downwind= 100 kts, base= 70 kts, final= 65 kts] and the participants' observed airspeed, kts) and vertical speed (the average difference between the reference optimal vertical speed [final= - 325 kts] and the participants' observed vertical speed, fpm).

In line with previous work, gaze behaviour was examined using traditional gaze measures, as well as static and dynamic entropy-based analyses (Ayala et al., 2022; Ayala et al., 2023b; Ziv 2016). Traditional gaze-based analysis was completed using the ten AOIs (Figure 6.3) that were discretized during pre-processing using a custom script. Specifically, we examined dwell time, average dwell duration, and dwell rate across all AOIs. Dwell time was defined as the total duration spent within a given AOI as function of total flight time, reported here as a percentage (i.e., with respect to total time). Average dwell duration was defined as the average duration of time (msec) of all uninterrupted dwells within a given AOI. Dwell rate was defined as the number of fixations that occurred within a given AOI over a given period of time (i.e., dwells/sec). Lastly, blink rate was defined as the number of blinks that occurred over a given period of time (i.e., count/sec).

The static entropy-based analysis was completed using the ten AOIs (Figure 6.3) that were discretized during pre-processing. Eye fixations in the ten AOIs were assigned a number from 1 to 10 indicating the AOI where the eyes fixated. A sequence of fixation locations was

then generated for each trial. Custom scripts were written in Python to compute both SGE and GTE (Ayala et al., 2021; Ayala et al., 2023b), which were then normalized (**Equation 1**) (Shannon, 1948).

$$H_{NORMAL} = H/H_{MAX}$$

Equation 1

SGE was computed by first producing a vector, $\mathbf{V}$, of length 10, where $\mathbf{V}_i$ was the total number of fixations at AOI $\mathbf{i}$. $\mathbf{V}$ was then divided by the total number of fixations in the sequence, so that $\mathbf{V}_i$ was the probability of a fixation landing at AOI $\mathbf{i}$. The probability vector $\mathbf{V}$ was then applied to **Equation 2** (Shannon, 1948).

$$H_{SGE}(V) = - \sum_{v \in V} v \cdot \log(v)$$

Equation 2

GTE was computed by first creating a 10x10 transition matrix, $\mathbf{M}$, where $\mathbf{M}_{i,j}$ was the total number of transitions from AOI $\mathbf{i}$ to AOI $\mathbf{j}$. Each row, $\mathbf{M}_{i,*}$, was divided by the sum of row $\mathbf{i}$, so that $\mathbf{M}_{i,*}$ represented the probability of fixation transition from AOI $\mathbf{i}$ to any of the ten AOIs. Finally, GTE was computed using **Equation 3** (Ciuperca and Girardin, 2007), applying the transition matrix $\mathbf{M}$ and the probability vector $\mathbf{V}$.

$$H_{GTE}(M) = - \sum_{i=1}^{10} V_i \sum_{j=1}^{10} M_{i,j} \cdot \log(M_{i,j})$$

Equation 3

The dynamic entropy-based analysis developed by Ayala et al. (2023) was completed using a 10 second average sliding window to show the probability of a fixation being inside the cockpit over the length of the trial, P(inside). The choice of a 10 second sliding window was

based on previous work that initially arbitrarily chose a 30 second sliding window but found it did not have sufficient resolution in detecting the observed characteristic cycling of gaze behaviour (Ayala et al., 2023b). Notably, this ‘gaze tunneling bout analysis’ is specifically different from “tunnel vision” in that we are not assessing if there is a loss/reduction of useful peripheral vision. Instead, gaze tunneling is particularly interested in quantifying the reduction in gaze transitions from the external environment to the cockpit gauges; a behaviour which has been tied to reduced pilot performance as a result of poor instrument scanning behaviour (Allsop and Gray 2014; Allsop et al., 2017; Xion et al., 2016). Nevertheless, to reduce the potential for confusion between tunnel vision and gaze tunneling, we have chosen to refer to this dynamic entropy-based analysis from here on out as cognitive tunneling bout analysis. Cognitive tunneling is a phenomenon in which an individual jeopardizes their ability to perceive all pertinent information in their environment because of a tendency to focus on a singular AOI (Bell et al., 2005; Engström et al., 2005; van Leeuwen et al., 2015). As such, it more accurately reflects the gaze behaviour that the dynamic-entropy based analysis captures.

To examine this phenomenon dynamically, Ayala and colleagues assigned a binary number based on whether the fixation was inside or outside the cockpit. $P(\text{inside})$ was then computed as the number of fixations inside the cockpit divided by the total number of fixations in the 10 second window. When $P(\text{inside})$ was equal to 1, the participant was continuously fixating inside the cockpit for at least 10 seconds. When $P(\text{inside})$ was equal to 0, the participant was continuously fixating on the outside scenery for at least 10 seconds (i.e., cognitive tunneling). This was employed to objectively monitor the temporal dynamics of gaze behaviour, with a focus on how attention was deployed inside and outside the cockpit. Cognitive tunneling is an important behaviour to quantify in pilot scanning. In this context, it refers to the absence of

pilot scanning toward other AOIs that have pertinent information regarding aircraft control. As such, it may impact pilot performance as well as situation awareness (Allsop and Gray, 2014; Allsop et al., 2017; Ayala et al., 2023b; Bell et al., 2005; Xion et al., 2016). A cognitive tunneling ‘bout’ was defined as a period of time in which fixations remained entirely outside of the cockpit for at least 10 seconds. These bouts were detected as connected components (subsequent values) of zeros in the probability time series. Number of bouts was defined as the number of instances a cognitive tunneling bout was detected within a trial. Bout duration (sec) was defined as the average duration of all bouts that occurred in a trial. Total bout time (sec) was defined as the sum of all bout durations to quantify the total time individuals demonstrated gaze behaviours reflective of cognitive tunneling.

Situation awareness was assessed using a subjective questionnaire and a scenario probe. The SART questionnaire (Taylor, 1990) is a post-trial self-report questionnaire that uses a 7-point Likert scale (1=Low; 7=High) across 10 dimensions of situation awareness. Note that this is collapsed into three larger dimensions of attentional demands, attentional supply, and situation understanding. These ratings are then combined to calculate a measure of situation awareness (SA).

$$SA = Understanding - (Demand - Supply)$$

Equation 4

6.2.6 Statistical Analysis

Aircraft control (i.e., airspeed), traditional gaze measures, and static entropy measures for successful trials were analyzed using a two-way repeated measures ANOVA with Phase of Flight (Downwind, Base, Final) and Task Difficulty (Easy, Difficult) as the independent variables. This specific analysis provides in-depth insight into the effect of increasing task

difficulty across various stages of flight. Additionally, landing performance, aircraft control (i.e., vertical speed), dynamic entropy measures and subjective situation awareness scores for successful trials were analyzed using a one-way repeated measures ANOVA with Task Difficulty (Easy, Difficult) as the only independent variable. Note that the vertical speed parameter of aircraft control was included in this ANOVA model as this was only measured for the final approach phase of flight. All ANOVAs were performed with an alpha level set at 0.05. The Bonferroni post hoc correction for multiple comparisons was applied for all post hoc analyses following the repeated measure ANOVAs to determine significant differences between variables. A secondary analysis was conducted using a linear mixed model to assess the effect of flight hours (between-subject expertise measure) on gaze behaviour, situation awareness, and performance measures, with Task Difficulty as the repeated measures variable. The linear mixed model analysis was also completed to examine the effect of gaze behaviour on situation awareness and performance measures.

6.3 RESULTS

Participant demographics with respect to flight hours showed that on average participants had 180 flight hours (SD= 75) with a minimum of 51 hours and a maximum of 280 hours.

6.3.1 The effect of task difficulty on landing performance and aircraft control

Landing performance. Completion time (sec) produced a main effect of task difficulty, $F(1, 23) = 294.054, p < 0.001, \eta_p^2 = 0.927$. Easy trials were completed significantly quicker ($X = 176.36$ sec, $SD = 25.02$) than difficult trials ($X = 243.98$ sec, $SD = 37.04$) (**Figure 6.4A**). Landing

hardness produced a main effect of task difficulty, $F(1, 23)= 11.594$, $p=0.003$, $\eta_p^2= 0.345$.

Specifically, easy trials were associated with significantly increased landing hardness (fpm) ($X= -118.12$ fpm, $SD= 53.48$) compared to difficult trials ($X= -87.72$ fpm, $SD= 31.53$) (**Figure 6.4B**).

Lastly, landing accuracy did not significantly change between task difficulty, $F(1, 23)= 1.104$, $p=0.304$ (**Figure 6.4C**).

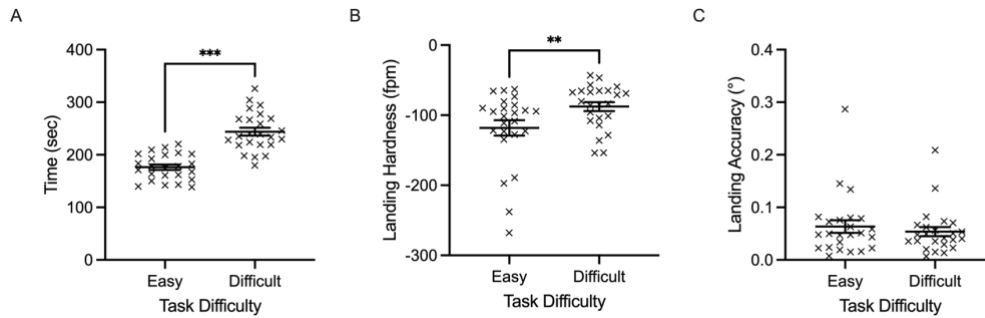


Figure 6.4. Individual data points and their respective group means for completion time (A), landing hardness (B), and landing accuracy (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Aircraft control. Mean airspeed (kts) demonstrated a main effect of phase of flight, $F(2, 46)= 519.682$, $p < 0.001$, $\eta_p^2= 0.958$, task difficulty, $F(2, 46)= 33.877$, $p < 0.001$, $\eta_p^2= 0.596$, and an interaction involving phase of flight and task difficulty, $F(2, 46)= 46.665$, $p < 0.001$, $\eta_p^2= 0.670$. Mean airspeed decreased significantly across each stage of flight (downwind: $X=93.52$ kts, $SD= 3.58$; base: $X=77.55$ kts, $SD= 5.93$; final: $X=62.69$ kts, $SD=2.50$). Moreover, difficult trials were associated with an increase in mean airspeed ($X= 79.41$ kts, $SD= 3.82$) compared to easy trials ($X= 76.43$ kts, $SD= 3.17$). However, post-hoc comparison of the interaction indicates that the task difficulty effect was only significant during the final approach phase of flight (easy: $X= 58.71$ kts, $SD= 2.66$; difficult: $X= 66.66$ kts, $SD= 3.28$). Airspeed variability (kts) also

yielded a main effect of flight phase, $F(2, 46) = 16.880$, $p < 0.001$, $\eta_p^2 = 0.423$, task difficulty, $F(2, 46) = 99.273$, $p < 0.001$, $\eta_p^2 = 0.812$, and an interaction involving flight phase and task difficulty, $F(2, 46) = 10.917$, $p < 0.001$, $\eta_p^2 = 0.322$. Specifically, airspeed variability increased significantly across each stage of flight (downwind: $X = 3.39$ kts, $SD = 1.62$; base: $X = 5.98$ kts, $SD = 2.79$; final: $X = 12.94$ kts, $SD = 1.67$). Easy trials had significantly higher airspeed variability compared to difficult trials (easy: $X = 8.84$ kts, $SD = 0.98$; difficult: $X = 6.04$ kts, $SD = 0.65$). However, the phase of flight by task difficulty interaction indicated that this was specific to the final stage of flight (easy: $X = 16.80$ kts, $SD = 2.12$; difficult: $X = 9.07$ kts, $SD = 1.48$). Lastly, airspeed RMSE demonstrated a main effect of phase of flight, $F(2, 46) = 13.123$, $p = 0.001$, $\eta_p^2 = 0.363$, task difficulty, $F(2, 46) = 107.577$, $p < 0.001$, $\eta_p^2 = 0.824$, and an interaction involving phase of flight and task difficulty, $F(2, 46) = 42.813$, $p < 0.001$, $\eta_p^2 = 0.651$. **Figure 6.5** illustrates a significant increase in airspeed RMSE across each phase of flight (downwind: $X = 7.89$ kts, $SD = 3.49$; base: $X = 10.69$ kts, $SD = 5.09$; final: $X = 14.03$ kts, $SD = 1.90$). Though the main effect of task difficulty suggests easy trials had increased airspeed RMSE ($X = 12.52$ kts, $SD = 1.88$) compared to difficult trials ($X = 9.23$ kts, $SD = 1.61$), the interaction shows that this is only the case during the final approach stage of flight (easy: $X = 18.22$ kts, $SD = 2.25$; difficult: $X = 9.84$ kts, $SD = 2.02$) (**Figure 6.5**).

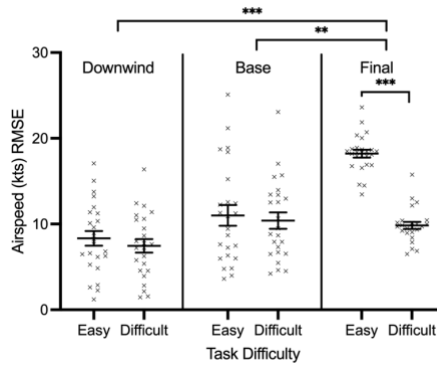


Figure 6.5. Individual data points and their respective group means for average airspeed RMSE are demonstrated for easy and difficult conditions across all downwind, base, and final approach stages of flight. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Mean vertical speed (fpm) yielded a main effect of task difficulty during the final approach phase of flight, $F(1, 23) = 146.312$, $p < 0.001$, $\eta_p^2 = 0.864$. **Figure 6.6A** demonstrates how difficult trials are associated with significantly lower mean vertical speed ($X = -285.85$ fpm, $SD = 41.85$) compared to easy trials ($X = -451.00$ fpm, $SD = 87.12$). Vertical speed variability (fpm) also demonstrated a main effect of task difficulty during the final approach phase of flight, $F(1, 23) = 124.816$, $p < 0.001$, $\eta_p^2 = 0.844$. Specifically, vertical speed variability was significantly higher during easy trials ($X = 246.15$ fpm, $SD = 48.14$) compared to difficult trials ($X = 139.52$ fpm, $SD = 37.36$). Vertical Speed RMSE demonstrated a main effect of task difficulty, $F(1, 23) = 65.785$, $p < 0.001$, $\eta_p^2 = 0.741$. **Figure 6.6B** illustrates how difficult trials are associated with a reduced RMSE ($X = 154.43$ fpm, $SD = 7.18$) compared to easy trials ($X = 286.53$ fpm, $SD = 15.88$).

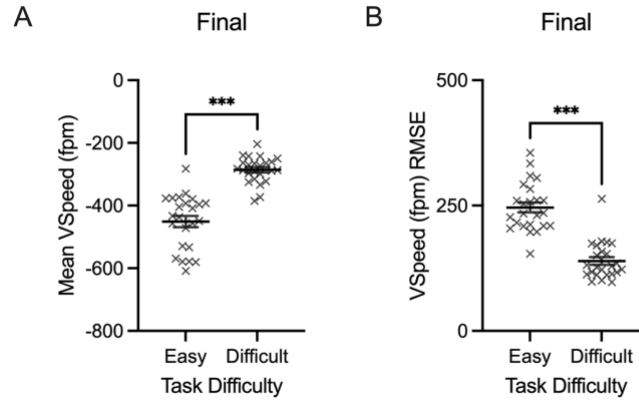


Figure 6.6. Individual data points and their respective group means for vertical speed (VSpeed) (kts) (A), and VSpeed root mean square error (RMSE) (B) are demonstrated for easy and difficult conditions across the final approach stage of flight. Error bars represent SEM.

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

6.3.2 The effects of task difficulty on gaze behaviour

Traditional Gaze Metrics. Dwell time (%) means and standard deviations for all AOIs across all task conditions are reported in **Table 6.1**. Dwell time (%) revealed significant changes across several AOIs associated with flight phase and task difficulty (Figure 6.7). Dwell time on the airspeed AOI demonstrated a main effect of flight phase, $F(2,46) = 12.881$, $p < 0.001$, $\eta_p^2 = 0.359$, and task difficulty, $F(1, 23) = 4.693$, $p = 0.041$, $\eta_p^2 = 0.169$. Specifically, airspeed AOI dwell time decreased on average in the difficult trials (-0.78%) relative to the easy trials. Airspeed AOI dwell time was at its highest during the base phase of flight, followed by the final, and downwind phases of flight. Attitude AOI demonstrated a main effect of phase, $F(2,46) = 32.560$, $p < 0.001$, $\eta_p^2 = 0.586$. Specifically, dwell time for the attitude gauge was at its highest during the base leg of flight, followed by the downwind, and final phases of flight. There was a

main effect of flight phase for the Altimeter AOI, $F(2, 46)= 46.605$, $p<0.0001$, $\eta_p^2=0.670$ and, $ps<0.001$, and for the power AOI, $F(2, 46)= 40.179$, $p<0.0001$, $\eta_p^2=0.636$. Dwell times on these AOIs decreased significantly from the downwind to the base and then final phases of flight.

VOR AOI showed a main effect of flight phase, $F(2, 46)= 7.448$, $p=0.007$, $\eta_p^2=0.245$, as well as an interaction between phase and task difficulty, $F(2, 46)=5.829$, $p=0.006$, $\eta_p^2=0.202$.

Decomposition of the interaction revealed that the difficult condition was associated with a longer dwell time (+1.79%) on the VOR AOI during the base phase of flight, $t(23)=-2.612$, $p=0.016$. Dwell times on the heading and vertical speed AOIs revealed a main effect of phase $F(2, 46)=6.093$ and 9.298 , $p=0.010$ and 0.003 , $\eta_p^2=0.209$ and 0.288 , respectively. Dwell times on both AOIs decreased from the initial downwind phase to the base phase, and again in the final phase of flight. Front window AOI dwell time demonstrated a main effect of phase, $F(2, 46)=604.499$, $p<0.001$, $\eta_p^2=0.963$. Dwell time through the front window increased significantly during the final phase of flight compared to both the downwind and base phases of flight. Left window AOI dwell time demonstrated a main effect of phase, $F(2, 46)=109.765$, $p<0.001$, $\eta_p^2=0.827$, as well as a main effect of task difficulty, $F(1, 23)=6.211$, $p=0.020$, $\eta_p^2=0.213$.

Specifically, dwell time on left window AOI was longest during the base phase of flight, which was followed by the downwind phase and the final approach phase of flight. Left window AOI dwell time also decreased significantly in the difficult condition (-0.94%) compared to the easy condition. Right window AOI dwell time did not show any significant changes due to flight phase or task difficulty. The significant changes in the distribution of attention (i.e., dwell time %) observed between easy and difficult conditions across all phases of flight are illustrated in **Figure 6.7**.

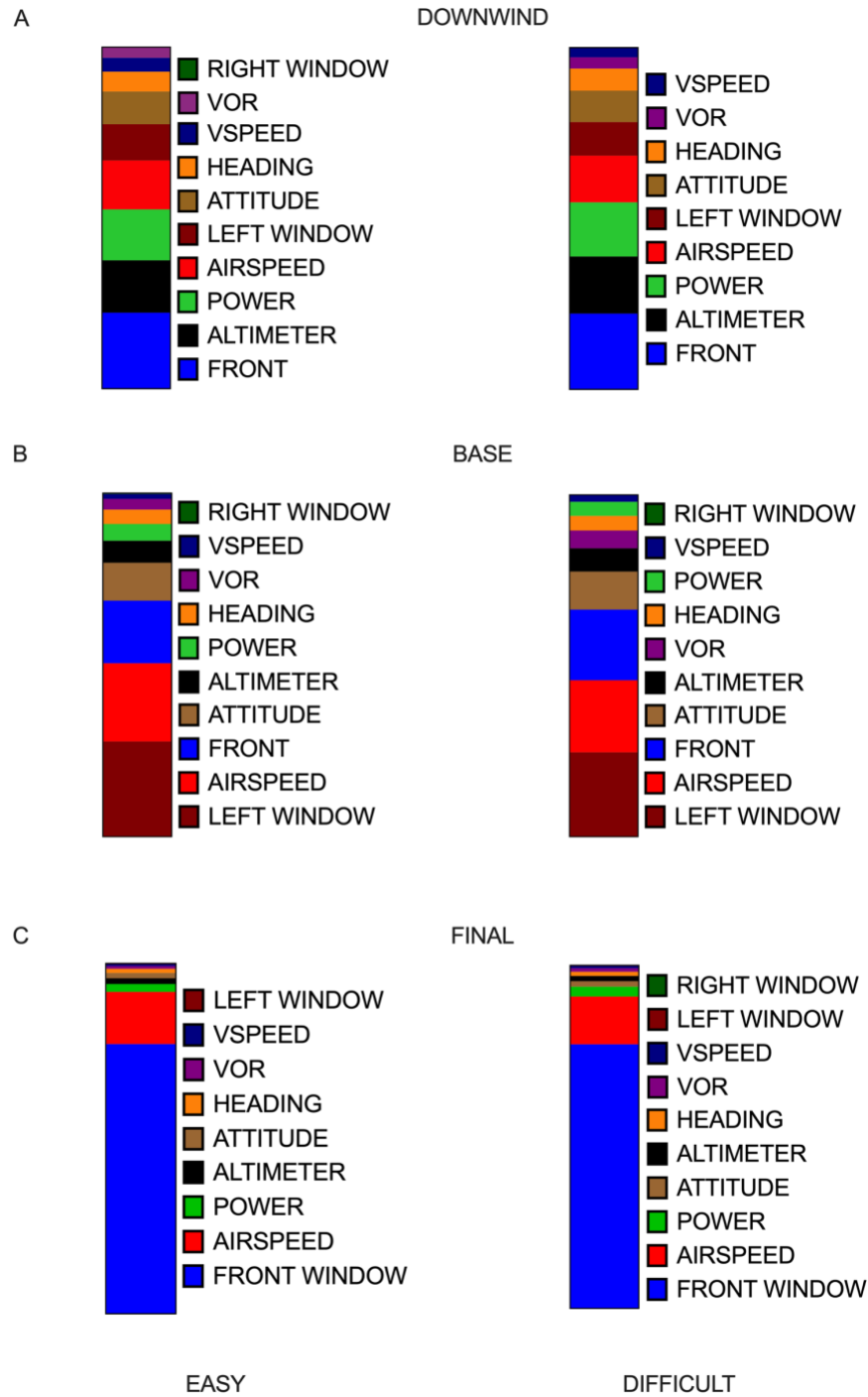


Figure 6.7. Candy-bar plots illustrate the overall dwell time (%) allotments across all AOIs during their respective phases of flight (A, B, C) for easy (left panels) and difficult (right panels) conditions.

Table 1. Dwell time (%) values calculated for all areas of interest for easy and difficult conditions across all phases of flight.						
Phase of Flight	Downwind		Base		Final	
Task Difficulty	Easy	Difficult	Easy	Difficult	Easy	Difficult
AIRSPPEED	10.22 (6.19)*	9.82 (6.64)*	18.92 (9.76)*	17.69 (7.63)*	13.36 (5.49)*	12.62 (4.75)*
ATTITUDE	6.85 (4.93)	6.62 (4.33)	9.16 (7.31)	9.35 (6.25)	1.38 (1.16)	1.44 (1.51)
ALTIMETER	10.95 (6.47)	12.02 (6.60)	5.33 (3.25)	5.72 (3.41)	1.47 (0.80)	1.40 (1.04)
VOR	2.28 (2.25)	2.39 (2.91)	2.59 (4.65)	4.38 (5.21)	0.69 (0.97)	0.92 (1.26)
HEADING	4.15 (5.98)	4.64 (5.89)	3.48 (4.53)*	3.59 (4.16)*	1.07 (1.77)	1.11 (1.06)
VERTICAL SPEED	2.89 (4.19)	2.03 (1.87)	1.17 (1.23)	1.59 (1.71)	0.46 (0.58)	0.57 (0.38)
POWER	10.66 (6.66)	11.40 (6.41)	4.04 (2.43)	3.45 (2.29)	2.00 (1.42)	2.60 (1.64)
FRONT WINDOW	15.90 (10.53)	15.96 (10.16)	15.12 (6.99)	17.36 (7.74)	69.16 (8.77)	69.84 (8.83)
LEFT WINDOW	7.59 (4.97)*	7.08 (5.85)*	23.02 (9.36)*	20.74 (7.59)*	0.21 (0.64)*	0.16 (0.22)*
RIGHT WINDOW	0.01 (0.03)	0.00 (0.00)	0.23 (0.64)	0.11 (0.27)	0.00 (0.00)	0.04 (0.12)

Note. Mean (standard deviation) dwell time (%) values across all areas of interest and task difficulty levels (i.e., easy, difficult). Significant changes between easy and difficult conditions are reported via * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Dwell rate means and standard deviations for all AOIs across all task conditions are reported in **Table 6.2**. Dwell rate (dwells/sec) was shown to be modulated by phase of flight and task difficulty to varying extents across several AOIs. Airspeed dwell rate showed a main effect of phase, $F(2, 46)=5.891$, $p=0.005$, $\eta_p^2=0.204$. Dwell rate at the airspeed gauge was highest during the downwind phase of flight and decreased significantly in the base and final phases of flight. Attitude dwell rate also demonstrated a main effect of phase, $F(2, 46)=13.400$, $p<0.001$, $\eta_p^2=0.368$. Attitude dwell rate was highest during the final phase of flight, which was followed by the downwind phase and then the base phase of flight. Altimeter dwell rate demonstrated a main effect of phase, $F(2, 46)=14.324$, $p<0.001$, $\eta_p^2=0.384$. Dwell rate toward the altimeter gauge was highest during the final phase of flight, which was significantly different from the base and downwind phases of flight. Front window dwell rate demonstrated a main effect of phase, $F(2, 46)=97.324$, $p<0.001$, $\eta_p^2=0.809$, which was associated with a significant reduction in dwell rate from the downwind phase of flight to the base phase, then again from the base

phase to the final phase of flight. Notably, there was a phase by task difficulty interaction, $F(2, 46)=4.896$, $p=0.020$, $\eta_p^2=0.176$. Decomposition of the interaction did not reveal any significant differences across the conditions, but it is worth noting that front window dwell rate was slightly higher (+0.01 dwells/sec) in the difficult condition than the easy condition during the downwind phase. In contrast, all other phases of flight demonstrated lower front window dwell rates for the difficult condition compared to the easy condition. Left window dwell rate yielded a main effect of phase, $F(2, 46)=10.405$, $p=0.002$, $\eta_p^2=0.311$. The downwind phase of flight had the highest left window dwell rate, which was significantly lower in the base and final phases of flight. Right window dwell rate was the only AOI to demonstrate a main effect of task difficulty, $F(1, 23)=4.844$, $p=0.038$, $\eta_p^2=0.174$. Specifically, the difficult condition was associated with an increase in dwell rate toward the right window (+0.04 dwells/sec) compared to the easy condition. All other AOIs did not demonstrate any significant changes due to phase of flight or task difficulty ($ps>0.083$).

Table 2. Dwell rate (dwells/sec) values calculated for all areas of interest for easy and difficult conditions across all phases of flight.

Phase of Flight	Downwind		Base		Final	
Task Difficulty	Easy	Difficult	Easy	Difficult	Easy	Difficult
AIRSPED	0.15 (0.07)	0.16 (0.09)	0.12 (0.04)	0.12 (0.03)	0.13 (0.04)	0.13 (0.04)
ATTITUDE	0.23 (0.13)	0.23 (0.12)	0.14 (0.04)	0.16 (0.05)	0.34 (0.21)	0.31 (0.23)
ALTIMETER	0.16 (0.07)	0.15 (0.06)	0.15 (0.05)	0.15 (0.05)	0.24 (0.13)	0.21 (0.08)
VOR	0.22 (0.22)	0.17 (0.18)	0.22 (0.19)	0.21 (0.23)	0.11 (0.13)	0.19 (0.16)
HEADING	0.2 (0.13)	0.23 (0.17)	0.19 (0.09)	0.22 (0.13)	0.25 (0.23)	0.18 (0.09)
VERTICAL SPEED	0.17 (0.12)	0.23 (0.17)	0.26 (0.28)	0.19 (0.14)	0.25 (0.37)	0.31 (0.28)
POWER	0.13 (0.06)	0.12 (0.04)	0.14 (0.11)	0.15 (0.09)	0.15 (0.05)	0.15 (0.04)
FRONT WINDOW	0.14 (0.04)	0.15 (0.04)	0.10 (0.03)	0.09 (0.03)	0.06 (0.02)	0.05 (0.02)
LEFT WINDOW	0.28 (0.12)	0.28 (0.10)	0.15 (0.06)	0.15 (0.06)	0.17 (0.33)	0.12 (0.17)
RIGHT WINDOW	0.04 (0.21)*	0.00 (0.00)*	0.13 (0.34)*	0.13 (0.23)*	0.00 (0.00)*	0.16 (0.41)*

Note. Mean (standard deviation) dwell rate (dwells/sec) values across all areas of interest and task difficulty levels (i.e., easy, difficult). Significant changes between easy and difficult conditions are reported via * $p\leq0.05$, ** $p\leq0.01$, *** $p\leq0.001$.

Dwell duration means and standard deviations for all AOIs across all task conditions are reported in **Table 6.3**. Average dwell duration demonstrated main effect of flight phase that was

specifically shown for a limited number of AOIs including, Attitude, $F(2, 46)=22.714$, $p<0.001$, $\eta_p^2=0.497$, Altimeter, $F(2, 46)=10.820$, $p<0.001$, $\eta_p^2=0.320$, Power, $F(2, 46)=8.598$, $p=0.001$, $\eta_p^2=0.290$, and the Front window, $F(2, 46)=58.488$, $p<0.001$, $\eta_p^2=0.718$. Attitude average dwell durations were longest during the base phase of flight, which was significantly longer than both the downwind and final approach phases of flight. Altimeter dwell durations were longest during the downwind phase of flight, then significantly decreased during the final phase of flight. Power dwell durations were longest during the downwind phase of flight and significantly decreased across each subsequent stages of flight (i.e., base and final approach). Front window dwell durations were longest during the final approach stage of flight, which became significantly shorter during each preceding phase of flight (i.e., base and downwind). All other AOIs did not demonstrate any significant changes as a result of phase of flight or task difficulty ($ps>0.05$).

Table 3. Average dwell duration (msec) values calculated for all areas of interest across all phases of flight.			
Phase of Flight	Downwind	Base	Final
AIRSPEED	532.91 (204.69)	565.06 (147.07)	524.47 (131.64)
ATTITUDE	480.56 (168.65)##*	453.15 (147.53)##^^^	349.93 (97.51)*^^^
ALTITUDE	361.47 (124.35)**	447.37 (101.27)^	280.25 (113.79)**^
VOR	413.65 (238.48)	418.71 (125.04)	380.76 (191.25)
HEADING	397.02 (183.28)	434.37 (166.41)	349.71 (146.11)
VERTICAL SPEED	343.73 (207.28)	413.82 (354.28)	311.95 (153.20)
POWER	650.19 (235.11)***	556.47 (151.79)^	453.59 (86.69)***^
FRONT WINDOW	491.99 (126.69)####	698.11 (198.73)####^^^	1349.43 (550.27)***^^^
LEFT WINDOW	240.07 (169.58)	402.44 (259.71)	244.93 (213.39)
RIGHT WINDOW	-	-	-

Note. Mean (standard deviation) average dwell duration (msec) values across all areas of interest and phases of flight (i.e., downwind, base, final). Significant changes between downwind and base phase of flight are reported via # $p\leq 0.05$, ## $p\leq 0.01$, ### $p\leq 0.001$. Significant changes between base and final phase of flight are reported via ^ $p\leq 0.05$, ^^ $p\leq 0.01$, ^^ $p\leq 0.001$. Significant changes between downwind and final phase of flight are reported via * $p\leq 0.05$, ** $p\leq 0.01$, *** $p\leq 0.001$.

Changes in blink rate (blinks/sec) across flight phases are shown in figure 6.8. Results demonstrated a main effect of flight phase $F(2, 46)=11.609, p=0.001, \eta_p^2=0.335$, and task difficulty, $F(1, 23)=4.955, p=0.036, \eta_p^2=0.177$, as well as an interaction between phase and task difficulty $F(2, 46)=3.215, p=0.049, \eta_p^2=0.123$ (**Figure 6.8**). The base phase of flight was associated with the highest blink rate (0.62 blinks/sec, SD= 0.47) compared to the downwind (0.25 blinks/sec, SD= 0.15) and final approach phases of flight (0.28 blinks/sec, SD= 0.19). Additionally, blink rate was significantly lower during the difficult condition (0.37 blinks/sec, SD= 0.17), compared to the easy condition (0.39 blinks/sec, SD= 0.17). Decomposition of the interaction revealed that blink rate was significantly lower during the difficult condition, specifically during the final phase of flight (easy: 0.29 blinks/sec, SD= 0.07; difficult: 0.26 blinks/sec, SD= 0.07), $t(23)=2.245, p=0.035$.

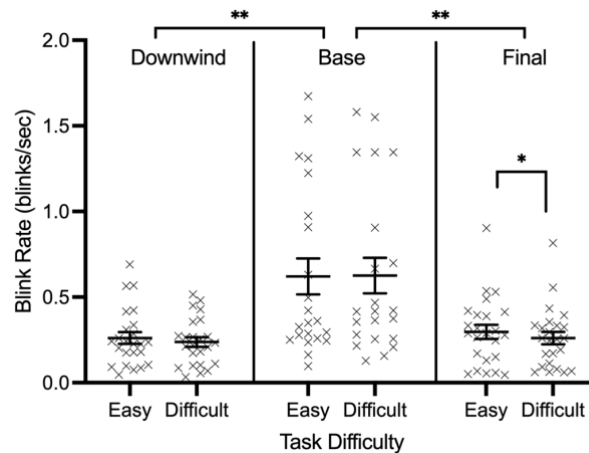


Figure 6.8. Individual data points and their respective group means for blink rate (blinks/sec) are demonstrated for each phase of flight. Error bars represent SEM. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

SGE and GTE means and standard deviations for all AOIs across all task conditions are reported in **Table 6.4**. SGE demonstrated a main effect of phase, $F(2, 46)=137.971$, $p<0.001$, $\eta_p^2=0.857$, task difficulty, $F(1, 23)=14.628$, $p=0.001$, $\eta_p^2=0.389$, and an interaction involving phase and task difficulty, $F(2, 46)=6.054$, $p=0.005$, $\eta_p^2=0.208$. SGE was highest during the initial downwind phase of flight and decreased significantly during the final approach stage. In general, the more difficult task condition was associated with greater fixation dispersion. However, decomposition of the phase by task difficulty interaction specifically revealed that SGE differed significantly between easy and difficult conditions (~ 0.19 bits) during the final approach phase of flight, $t(23)=-4.292$, $p<0.001$ (**Figure 6.9A**). GTE demonstrated similar main effects of flight phase, $F(2, 46)=33.413$, $p<0.001$, $\eta_p^2=0.592$, task difficulty, $F(1, 23)=19.401$, $p<0.001$, $\eta_p^2=0.458$, and an interaction involving phase and task difficulty, $F(2, 46)=15.943$, $p<0.001$, $\eta_p^2=0.409$. GTE was highest during the base phase of flight, which was significantly lower during the downwind phase of flight and the final approach stage of flight. The difficult condition was associated with a higher GTE. However, decomposition of the interaction involving phase and task difficulty showed that GTE was significantly different between the easy and difficult task conditions (~ 0.21 bits) during the final phase of flight only, $t(23)=-5.816$, $p<0.001$ (**Figure 6.9B**).

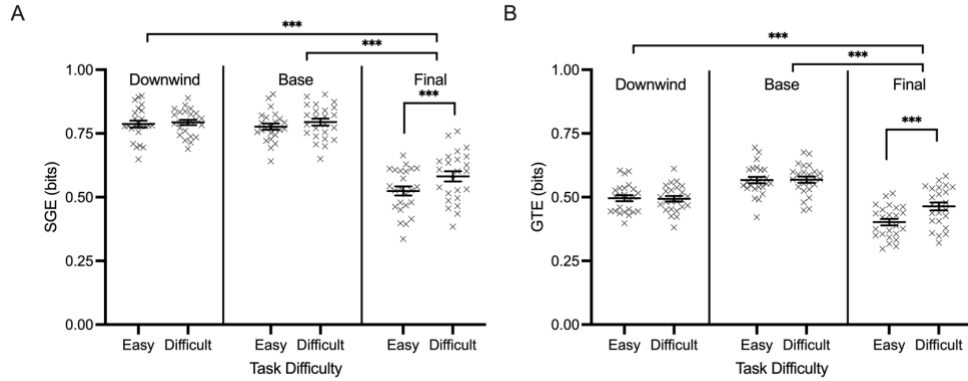


Figure 6.9. Normalized individual data points and their respective group means for stationary gaze entropy (SGE) (bits) (A) and gaze transition entropy (GTE) (B) are demonstrated for easy and difficult conditions each phase of flight (i.e., downwind, base, final approach) . Error bars represent SEM. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

Phase of Flight	Downwind		Base		Final	
Task Difficulty	Easy	Difficult	Easy	Difficult	Easy	Difficult
Stationary Gaze Entropy (SGE)	2.61 (1.07)	2.6 (0.83)	2.58 (0.92)	2.64 (1.08)	1.74 (1.42)***	1.93 (1.61)***
Gaze Transition Entropy (GTE)	1.65 (0.93)	1.64 (0.83)	1.88 (0.98)	1.89 (0.93)	1.33 (1.02)***	1.54 (1.22)***

Note. Mean (standard deviation) entropy (bits) values across all phases of flight (i.e., downwind, base, final) and task difficulty levels (i.e., easy, difficult). Significant changes between easy and difficult conditions are reported via $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

Dynamic gaze behaviour. Number of cognitive tunneling bouts revealed a main effect of task difficulty, $F(1, 23)=10.360$, $p=0.004$, $\eta_p^2=0.311$. Specifically, number of bouts was significantly higher in the difficult condition ($X=1.69$, $SD= 1.24$) compared to the easy condition ($X=1.07$, $SD= 0.43$) (**Figure 6.10A**). Total bout time (sec) also demonstrated a main effect of task difficulty, $F(1, 23)=8.006$, $p=0.010$, $\eta_p^2=0.258$. The difficult condition was associated with significantly longer total bout time ($X=27.91$, $SD= 21.96$), compared to the easy condition

($X=18.41$, $SD= 8.39$) (**Figure 6.10C**). Average bout duration was not significantly modulated by task difficulty ($p=0.275$) (**Figure 6.10B**).

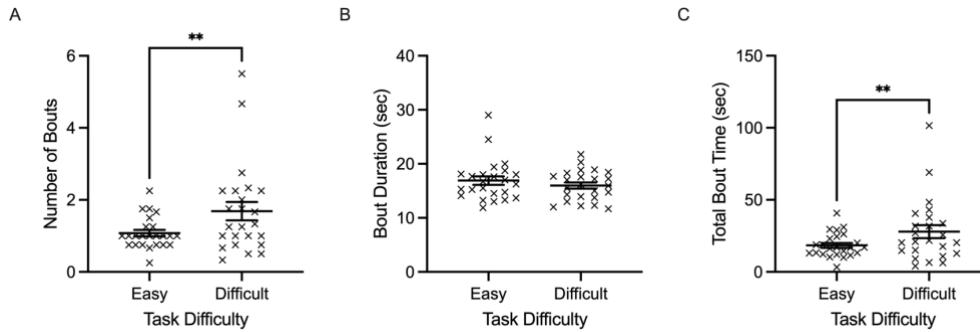


Figure 6.10. Individual data points and their respective group means for number of cognitive tunneling bouts (A) average bout duration (B), and total cognitive tunneling bout time (C) are demonstrated for easy and difficult conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

6.3.3 The effects of task difficulty on situation awareness

Situation awareness. Subjective scores from the SART questionnaire produced a general SA score, which demonstrated a main effect of task difficulty, $F(1, 23)= 22.769$, $p < 0.001$, $\eta_p^2 = 0.497$. Specifically, subjective SA scores were lower for difficult trials ($X= 17.76$, $SD= 6.50$) compared to easy trials ($X= 21.19$ $SD= 5.23$) (**Figure 6.11A**). A closer examination of the SART questionnaire subcomponents revealed that SA demand and SA supply also yielded a main effect of task difficulty, $F(1, 23)= 57.280$ and 13.931 , $ps < 0.001$, $\eta_p^2 = 0.714$ and $.377$, respectively.

Figure 6.11C and D shows how both SA demand and SA supply components increased in the difficult condition (SA demand: $X= 11.96$, $SD= 7.35$; SA supply: $X= 19.66$, $SD= 3.29$) compared to the easy condition (SA demand: $X= 7.57$, $SD= 6.27$ SA supply: $X= 18.33$, $SD=$

3.5). The SA understanding component did not reveal a main effect of task difficulty ($p=0.140$) (Figure 6.11B).

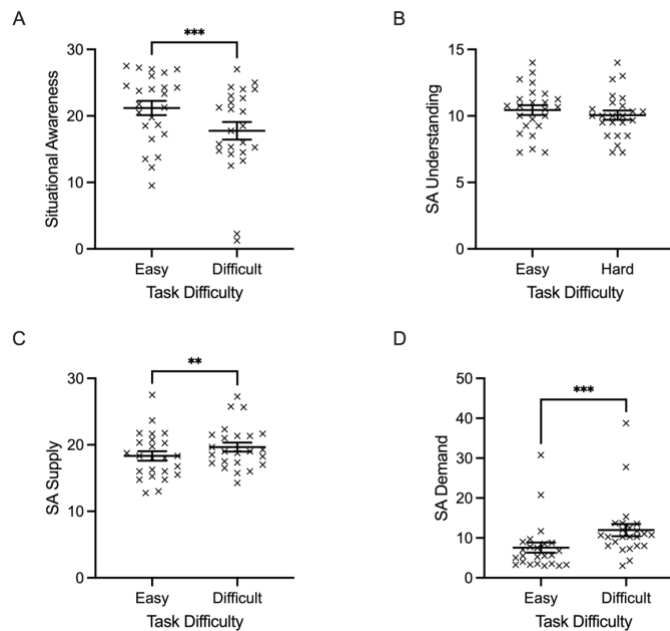


Figure 6.11. Individual data points and their respective group means for Situation Awareness (SA) Score (A), SA Understanding subcomponent score (B), SA Supply subcomponent score (C), and SA Demand subcomponent score (D) are demonstrated for easy and difficult conditions. Error bars represent SEM. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

6.3.4 Regression Analyses

A linear mixed model analysis was run in R Studios (version 4.3.1) to examine the relationship between the developed dynamic gaze metrics (i.e., number of cognitive tunneling bouts and total bout time) and pilot flight experience (i.e., flight hours) (Table 6.5). In the first model, we included the number of bouts as the dependent variable, task difficulty as a fixed effect, flight hours as a random effect, and the interaction between flight hours and task

difficulty. There was a main effect of task difficulty, $F(1, 46)=17.29, p<0.001, \eta_p^2=0.44$, flight hours, $F(1, 46)= 4.71, p= 0.041, \eta_p^2=0.18$, and an interaction between flight hours and task difficulty, $F(1, 22)=8.88, p=0.006, \eta_p^2=0.29$ (**Table 6.5**). Specifically, the slopes were significantly different between the easy and difficult conditions when compared between the same group of pilots, with more cognitive tunneling bouts being observed in pilots with lower hours during the difficult condition (**Figure 6.12A**).

Dependent Variable	Task Difficulty		Flight Hours		Task Difficulty x Flight Hours Interaction	
	Coefficient ,[95%CI]	R ²	Coefficient ,[95%CI]	R ²	Coefficient ,[95%CI]	R ²
Number of Bouts	0.432, [0.067, 1.155]	0.081	-0.005 , [-0.008, -0.001]	0.1	Easy: -0.00125, [-0.006, 0.004] Difficult: -0.00804, [-0.013, -0.003]	0.68
Total Bout Time	6.722, [-0.153, 19.167]	0.059	-0.075 , [-0.141, -0.008]	0.081	Easy: -0.0199, [-0.109, 0.069] Difficult: -0.1294, [-0.219, -0.039]	0.65

Note. Estimated marginal (EM) mean of linear trend characteristics for the number of bouts and total bout time models are provided including their respective coefficients (β), 95% lower/upper confidence levels, and R-squared values.

In the second model, we included total bout time as the dependent variable, task difficulty as a fixed effect, flight hours as a random effect, and the interaction between flight hours and task difficulty (**Table 6.5**). Each subject was also included as a random effect. There was a main effect of task difficulty, $F(1, 46)= 13.12, p=0.001, \eta_p^2=0.37$, and an interaction between flight hours and task difficulty, $F(1, 22)= 6.89, p=0.015, \eta_p^2=0.24$ (**Table 6.5**). Specifically, the slopes were significantly different between the easy and difficult conditions when compared between the same group of pilots, with longer total gaze time being observed in pilots with lower hours during the difficult condition (**Figure 6.12B**). Note that regression analyses involving all other dependent variables did not reach significance.

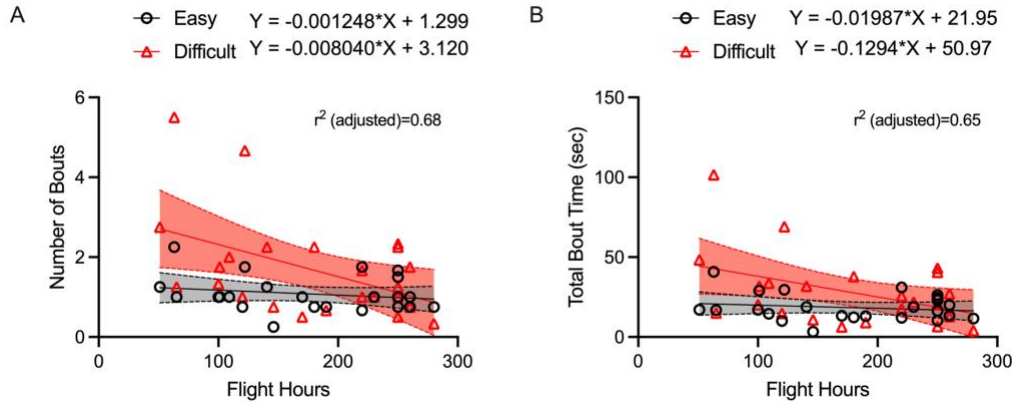


Figure 6.12. Scatter plots demonstrate the relationship between the number of cognitive tunneling bouts and pilot flight hours (A), and the total cognitive tunneling bout time (sec) and pilot flight hours (B). Black circles indicate easy condition participant averages, red triangles indicate difficult condition participant averages. Lines indicate the trend for all subjects (i.e., line of best fit). Highlighted bands around line of best fit indicate the 95% confidence interval.

6.4 DISCUSSION

Aviation accidents related to human error have been increasingly associated with poor pilot monitoring as it can negatively impact the degree to which operators receive adequate information to understand and predict the changing circumstances; thus, directly impact performance and SA (Boeing, 2021; National Transportation Safety Board, 1994; Shappel and Wiegmann, 2000; Stanton et al., 2017). As such, the current study provides one of the first accounts for how gaze behaviour changes simultaneously with flight performance and SA throughout the various stages of landing (i.e., downwind, base, final approach) during different task difficulty conditions to gain insight into low-time pilot monitoring patterns. Participants were asked to perform the landing task in high visibility, visual flight rules (VFR) conditions that included a task difficulty manipulation based on the absence (easy) or presence (difficult) of

strong crosswinds. Several notable contributions emerged from this study. First, there was partial support of our initial hypothesis that gaze behaviour would reflect a greater need to allocate more time and attention toward fewer, task critical AOIs during the difficult condition. Specifically, a reduction in attention allocation toward fewer cockpit AOIs was coupled with an increase in cognitive tunneling behaviour during the difficult condition. Moreover, the allocation of attention across most AOIs was significantly impacted by flight phase. Second, an exploratory analysis that employed linear mixed model regressions found significant associations between dynamic gaze metrics and pilot flight hours. These contributions and their implications for integrating gaze behaviour analysis into pilot training and assessment are discussed below.

6.4.1 Examining the effectiveness of the task difficulty manipulation

Landing performance data provided some evidence that task difficulty was effectively manipulated using varying weather conditions (i.e., no wind versus significant crosswind); however, some results were contrary to what was expected. Similar to previous work (Ayala et al., 2023b; Diaz-Piedra et al., 2019), completion time was ~68 seconds (~39%) longer in the difficult condition compared to the easy condition. Although the difficult condition required more time to complete, task performance data demonstrated a speed-accuracy trade-off in that other flight performance parameters were improved. This contradicts what was reported in previous work (Ayala et al., 2023b), where more difficult scenarios were associated with longer completion times and increased landing error. However, it is important to recall that the current study recruited a more homogenous pilot cohort with more flying experience; and thus, they might have been able to handle the task difficulty manipulations used in previous work with less experienced participants (Ayala et al., 2023b). For instance, the current study demonstrated that the difficult condition was also associated with a reduction in landing hardness (~30 fpm),

airspeed RMSE and variability, as well as vertical speed RMSE and variability. This reduction in error and performance variability was specifically seen during the final approach stage of flight, which was shown to have the highest error and variability across all stages of landing (i.e., downwind, base, final approach). Notably, pilot performance during the easy condition was not necessarily bad. According to Transport Canada landing guidelines for PLL pilots, the observed landings were all considered successful and safe. Pilots simply performed better during difficult trials, likely because they were being more attentive to the task demands and performing the task with a higher degree of effort, particularly during final approach. Work by Stuhr and colleagues (2018) further supports this assertion as they demonstrated that the impact of cognitive control processes on motor skill proficiency depends on performance variability. In other words, when performance variability is high (i.e., typical in novel or highly complex/difficult motor tasks such as the final approach phase of flight in this study), individuals are more likely to engage cognitive control processes to assist in the successful performance of the task at hand.

Another potential account for these results could be that the final approach phase of flight was the only phase of flight that had 22 knots of headwind; a component that was not present in previous work (Ayala et al., 2023b). Headwind is notoriously known to make it easier for pilots to control the aircraft as it provides additional lift to the aircraft and at lower speeds (Federal Aviation Administration, 2021). This in turn would be associated with improved aircraft control and landing performance (i.e., reduced variability and deviation from pre-set parameters) **(Figures 6.4-6.6)**. Both accounts can explain the reported results. Accordingly, more research is required to fully understand how the performance parameters are being modulated by the task difficulty manipulation used in the current study. Specifically, future work could examine an alternative task difficulty manipulation that doesn't involve the introduction of winds, introduces

head winds during the final approach but no cross wind to the “easy” trial, or significantly reduces the headwind component encountered during the final approach phase of the landing scenario.

From the performance measures alone, it could be argued that the task difficulty manipulation did not necessarily increase the difficulty of the task itself, but rather made it more engaging. However, the additional cognitive resources devoted to maintaining/improving task performance were likely reallocated from other task-related aspects of flight performance, such as SA or changes in information processing (demonstrated through changes in gaze behaviour). Indeed, seminal research has shown that SA plays a critical role in pilot performance as it involves the degree to which operators *receive* (i.e., perceive and process) adequate information to *understand* the changing conditions in their environment, and *project* the impact on future circumstances (Endsley, 1995). This is a crucial aspect of aircraft control to consider, because even though flight performance may not be overtly different across various flight scenarios (i.e., changes in wind conditions), a pilot’s ability to receive incoming information, comprehend it, and project its impact on future circumstances may still be impaired, which will affect their ability to problem solve and make informed decisions should an unexpected event (i.e., in-flight emergency) occur. The current study demonstrated that subjective SA, as indicated through the SART, decreased by ~8% in the difficult condition relative to the easy condition (**Figure 6.11**). An in-depth assessment of the SART SA sub-components revealed that attentional supply (i.e., arousal, spare mental capacity, level of concentration, division of attention) increased by approximately 5% (**Figure 6.11**), which supports the earlier claim that participants were being more attentive to task demands and exerted more effort during the difficult condition. However,

this was substantially overshadowed by the ~21% increase in the attentional demand sub-component (i.e., how unstable, variable, and complex was the situation) (**Figure 6.11**).

In light of the reduction in aircraft control error and variability during difficult trials, the SA results are interesting in the sense that improved aircraft control was hypothesized to be associated with an increase in SA, whereas a reduction in SA was assumed to be associated with poorer aircraft control. Three alternative explanations are proposed to account for the reported findings where pilot performance improvements were seen in parallel with a reduction in SA. First, an argument could be made that subjective SA as measured through the SART may be more reflective of participant confidence level and not necessarily a true representation of SA (Endsley et al., 1998; Selcon et al., 1991). As such, the reduced SART scores may indicate that participants were less confident about their performance during the difficult condition as they had to devote more attentional resources toward the task relative to the easy condition. A second alternative explanation could be that if pilots were concentrating more on aircraft control during the difficult condition, perhaps this was at the cost of no longer attending to other things going on around them; thus, resulting in a reduction in SA. A third explanation could also argue for the existence of a ‘buffer’ effect that may be associated with pilot proficiency. In other words, a reduction in SA may be compensated with greater pilot proficiency when managing increased task demands, as they will have a lower chance of exceeding the pilots’ cognitive resources (Brams et al., 2018; Brams et al., 2020; Dehais et al., 2017). Since the current study recruited pilots who were already licensed, and thus experienced in flying the simulated aircraft under the tested flight conditions (i.e., little-no winds and high crosswind conditions), it is likely that they had the capacity to effectively devote additional cognitive resources to further support task performance in the presence of reduced SA during the difficult landing condition. The gaze

results (discussed below) suggest that the second and third explanations may be plausible accounts for the reported findings as there is evidence of a broad reduction in the scanning of environmental stimuli, but there is also data suggesting that some changes in gaze behaviour work to compensate for increases in task difficulty that might help prevent reductions in flight performance. Still, these assertions should be further tested in future research that significantly challenge the recruited pilot pool and examines how different gaze patterns align with how pilots manipulate aircraft control inputs (i.e., throttle, pitch, roll inputs).

In addition to the reported SA findings, several gaze variables were significantly modulated by increases in task difficulty. For instance, blink rate was reduced (**Figure 6.8**), the allocation of attention was selectively increased toward the front window while it decreased across a number of other less-relevant AOIs (**Figure 6.7**), and the frequency and total duration of cognitive tunneling gaze behaviours increased when task difficulty increased (**Figure 6.10**). Note that the dwell time findings reported here differ from earlier work (Ayala et al., 2023b), which reported an increase in front window dwell time during the difficult condition compared to the easy condition. The previous study suggested that this change in gaze behaviour reflected the need to allocate more attention toward fewer task critical AOIs; of which, the front window was a prime source for monitoring and extracting necessary information required to land during challenging wind conditions (Ayala et al., 2023b; Beall and Loomis, 1997; Di Nocera et al., 2007; Gray et al., 2014; Kim et al., 2008; Mertens, 1981; Sarter et al., 2007). In partial support for this expected finding, a comprehensive analysis of dwell time changes during each stage of flight demonstrated a reduction in attention toward the left window and airspeed indicators, but it failed to demonstrate a significant increase in dwell time toward the front window. This was further examined by aggregating the gaze data for a broader analysis of dwell time patterns

across the AOIs during difficult and easy scenarios without the segregation of data by flight phase. This additional analysis confirmed what was reported in earlier work showing a reduction in a number of AOIs (i.e., airspeed, attitude, altimeter, left window) and an associated increase in front window dwell time (~10%). Accordingly, we suggest that analysis of gaze data across various phases of flight is appropriate to assess information processing changes associated with the different sub-goals linked to each stage (Dehaise et al., 2020). However, the segregation of time-normalized dwell time data into stages of landing negatively impacts our ability to draw conclusions about how specific patterns of attention allocation across numerous AOIs- which have variable time courses- are impacted by task difficulty. It is also important to note that this limitation may be a consequence of a reduction in power for this specific analysis. Previous work examined task difficulty changes across all AOIs but did not further segregate the data by landing stage (Ayala et al., 2023b). As such, it may also be the case where more participants would have been required to properly replicate previous work with the additional phase of flight variable. Nevertheless, these shifts in gaze behaviour are proposed to reflect a necessary shift in top-down attentional controls imposed on visual scanning during challenging task demands to help focus attention to the appropriate object at the appropriate time (Ayala et al., 2023b; Ayala et al., 2022; Bellenskes et al., 1997; Brams et al., 2018; Eysenck et al., 2007; Shiferaw et al., 2019). Moreover, the blink rate findings corroborate the suggested relationship that blink rate depression is associated with an increase in task difficulty as the current findings demonstrated the lowest blink rate to coincide with the most challenging phase of landing (i.e., final approach) (Glaholt, 2014; Peißl et al., 2018; Ziv, 2016).

An interesting gaze outcome that was contrary to what was hypothesized was the significant increase in SGE and GTE during the difficult trials, particularly during the final

approach stage of landing. Previous work demonstrated that increases in task difficulty were associated with a reduction in SGE and GTE (Ayala et al., 2023b). Several reasons may help to explain the divergent results. First, it is important to remember that the cohort in the previous study consisted of ~56% ab-initio pilots with little to no flight experience while the current study exclusively enrolled licensed pilots. Conceivably, flight experience has a significant effect on gaze behaviour. For instance, licensed pilots may be more aware and capable of increasing their gaze dispersion and sequence complexity to gather more information in a more efficient manner to aid in the management of crosswind conditions (i.e., lateral slip/crabbing methods to enhance aircraft control) (Brams et al., 2018; Brams et al., 2020; Federal Aviation Administration, 2021). Second, it is possible that the pilots recruited in the current study were not challenged to the same extent as the participants recruited in previous work (Ayala et al., 2023b). As such, it could be hypothesized that the SGE and GTE would similarly decrease in response to a significantly more difficult condition if the current study provided a greater challenge. In this case, the hypothetical SGE and GTE reductions may also reflect a shift in top-down attentional control mechanisms to selectively allocate visual attention and focus visual scanning to highly critical AOIs when task difficulty is high (Ayala et al., 2023b). Future work should include a significantly more difficult task (e.g., dual-task paradigm) to stress the cognitive control resources of pilots during a simulated flight. Since the reported increase in gaze scanning distribution and sequence complexity (**Figure 6.9**) were seen in parallel with a reduction in aircraft control variability (**Figure 5 & 6**) during the final stage of flight for the difficult relative to the easy condition these particular gaze data are suggested to reflect a compensatory mechanism to support task performance (Kübler et al., 2015).

6.4.2 Flight phase parameters and experience influence information processing

The findings presented clearly demonstrate that all traditional gaze metrics (i.e., dwell time percentage, dwell rate, and average dwell duration across AOIs) vary significantly between flight phases, which most likely reflects the differing task sub-goals specific to each stage of flight (Ayala et al., 2023b; Badu et al., 2019; Dehais et al., 2021; Di Nocera et al., 2007; Glaholt, 2014; Gray et al., 2014; Kim et al., 2008; Sarter et al., 2007; Van de Merwe, 2012). For instance, during the downwind flight phase participants are normally completing their landing checklists and flows to configure the plane for landing (Federal Aviation Administration, 2021). This was manifested in the downwind gaze patterns which had the highest dwell times for the front window, altimeter, power, and airspeed indicators, which are the various gauges that require attention when configuring the plane to land (i.e., continue flying straight ahead [antiparallel to runway heading], reduce power to 15,000 rpm, maintain circuit altitude, and ensure airspeed is configured for this landing stage [100kts]). Moreover, further exploration of non-AOI data demonstrated that the greatest amount of time devoted to looking at the landing checklist and other configuration regions not included in the 10 pre-defined AOIs was during the downwind phase of flight (16%). Although this was not included for analysis in the current study (<5% of total dwell time), it provides additional support for how gaze behaviour supports flight phase-specific task goals. The base leg of flight was associated with the highest dwell times for the airspeed, attitude, and the left and front window AOIs. This was tied to the fact that pilots were required to complete the turn-to-base and turn-to final approaches during this leg of flight while maintaining the recommended flight speeds (Federal Aviation Administration, 2021). As such pilots frequently monitored aircraft roll via the attitude gauge, landmarks that helped determine when to begin and end the turn (i.e., runway visual via the left window), and the airspeed gauge

to adjust airspeed as needed in between stage transitions (i.e., downwind= 100 kts, base= 70 kts, final= 65 kts). Last, the main goals of the final approach stage of landing were to maintain the glide slope, center alignment with the runway, and transition the aircraft from a normal approach attitude to a landing attitude (i.e., flare) (Federal Aviation Administration, 2021). Visual cues that help monitor each of these goals are predominantly located through the front window and serve as the main reason why the allocation of attention to this AOI is greatest during the final approach (Ayala et al., 2023; Dehais et al., 2021; Di Nocera et al., 2007; Glaholt, 2014; Gray et al., 2014; Kim et al., 2008; Sarter et al., 2007). The airspeed gauge was also a critical AOI as final approach speed is important for maintaining optimal glide slope and preventing an engine stall. These findings were supported by similar, though less robust, trends in the dwell rate and average dwell duration data.

In line with the traditional gaze findings, more computationally complex measures of gaze behaviour including SGE and GTE demonstrated significant reductions from the downwind/base stages of landing to the final approach stage of landing (**Figure 6.9**). This further supports the claim that gaze behaviour changes are contingent on flight phase specific sub-goals. For example, higher SGE and GTE suggest increased dispersion and more gaze shifts across the AOIs that require continual monitoring to configure the flight during the downwind and base stages compared to the final approach stage of flight; wherein the focus of attention routinely becomes restricted to 2-3 AOIs. With respect to other dynamic gaze measures of information processing, the analysis of the moment-to-moment changes in gaze fixation and dispersion provided insight into broader changes pilot monitoring and cognitive tunneling (i.e., prolonged outside gaze fixation) behaviours over time which were influenced by the task difficulty manipulations, but more so dictated by flight experience (i.e., flying hours). In line with previous

work (Ayala et al., 2023b), the current study demonstrated an increase in cognitive tunneling bouts and total cognitive tunneling bout time in the difficult condition compared to the easy condition. In other words, the difficult condition was associated with more frequent periods of time where pilots neglected to scan their cockpit. Although this could have resulted in a tunneling of attention toward any other AOI, what was observed in previous studies (Allsop and Gray 2014; Allsop et al., 2017; Ayala et al., 2023b; Ayala et al., 2024; Xion et al., 2016) alongside current work was a fixation of attention outside. This is a critical aspect of gaze behaviour to examine as a lack of scanning during phase(s) of flight that require continuous monitoring of flight parameters (i.e., landing) and aircraft state via gauges inside the cockpit can result in disastrous consequences should an unexpected/hazardous event occur. This phenomenon was similarly reported in driving simulation studies that found that the higher proportion of dwell time allocated to the center of the road and a reduction in gaze dispersion were associated with driver distraction, reduced hazardous event detection, and increased driver cognitive load (Reimer, 2009; Wang et al., 2014; Yang et al., 2018; Yang et al., 2022). The increased cognitive tunneling findings seemingly contradict the increase in final approach SGE and GTE during difficult trials. However, it is important to remember that the dynamic bout analysis is examining duration-based cognitive tunneling tendencies over time while time-averaged entropy measures only consider the frequency and sequence of AOIs (Ciuperca and Girardin, 2007; Shannon, 1984; Shiferaw et al., 2019). Moreover, the cognitive tunneling bout analysis provides a marker of poor monitoring behaviour that is independent of flight stage and instead considers the duration of time that gaze is diverted from cockpit gauge monitoring across the entire trial whereas the SGE/GTE findings in question are demonstrated only for the final approach phase of flight (Ayala et al., 2023b; Xion et al., 2016; Ziv, 2016). In fact, when SGE

and GTE are averaged across the entire trial, there is no effect of task difficulty; a finding that suggests that although fixation dispersion and gaze sequence complexity don't change as a function of task difficulty in the recruited participant pool, the cognitive tunneling behaviour (i.e., cockpit monitoring behaviour) remains an informative measure of changes in information processing. Moreover, the current study makes an important contribution by further demonstrating how pilot experience affects cognitive tunneling behaviour. The linear mixed model regressions demonstrated that cognitive tunneling events account for up to 68% of variance in pilot flight hours; a finding that provides evidence for the claim that dynamic cognitive tunneling bout analysis is a robust measure of pilot monitoring behaviours that change with experience. This is significant because it provides the first gaze-base measure that can identify differences in pilot proficiency within a tightly defined recruitment pool (i.e., flight hours range: 51-280hrs). Furthermore, the bout regression analysis provides additional support for the extent to which low time pilots need to be challenged before one can observe meaningful changes in gaze behaviour as a function of flight hours. This is demonstrated by the task difficulty by cognitive tunneling (number and total time) interaction suggesting the relationship for the currently defined flight hour range is present, but only in the more difficult condition when the capabilities of pilots with fewer flying hours are being challenged.

Notably, the implementation of a number of gaze metrics examined in the current study (i.e., dwell time phase plots, SGE/GTE, and cognitive tunneling) into pilot training and assessment protocols could provide easily interpretable metrics that are objectively related to pilot proficiency levels, and identify instances where pilots demonstrate deviations from optimal monitoring behaviours. In example, dwell time plots may be used to assess where pilots are allocating their attention and if it aligns with task-relevant sources of information that require

visual processing, SGE/GTE plots can aid in identifying pilots who are deviating from pre-determined task norm values, and cognitive tunneling measures can be used to assess poor monitoring behaviours that directly align with International Air Transport Association (IATA) competency assessment questions about “How Many” and “How Often” (Guidance Material, 2023) these behaviours occur. These capabilities further enhance the extent to which gaze metrics, such as those described here, can be used as an assessment and training tool.

6.4.3 Limitations and future directions

The current study provides several novel insights into the way task difficulty and flight phase parameters impact task performance, gaze behaviour and SA. However, the current results are constrained by at least four methodological limitations. First, the tight pilot recruitment pool limited the extent to which the regression model analysis could relate flight experience (i.e., flying hours) to dynamic gaze measures (i.e., number of cognitive tunneling bouts and total cognitive tunneling bout time). Future studies seeking to gain a better understanding about the utility of the novel dynamic gaze measures (Ayala et al., 2023b) in characterizing pilot flight hours or SA should include a larger range of pilot experience backgrounds that extend from the early ab-initio level to the instructor pilot level. Second, the extent to which other relationships could be thoroughly assessed was also undermined by the small participant pool. For example, the ability to examine the relationship between dynamic gaze behaviours and performance or SA was limited by the fact the models demonstrated a lack of power. As such, no conclusion could be made about the utility in using cognitive tunneling bout analysis to ascertain pilot SA or task performance. This could be improved in future work by increasing the number of trials each participant completes, or by conducting the experiment in Instrument Flight Rules (IFR) conditions where dynamic gaze behaviour and/or SA would be stressed. A third related

limitation was that there was no singular measure of landing performance that could be used to effectively determine relationships between flight performance and SA or dynamic gaze metrics. As such, future work should involve instructor pilots that are familiar with the assessment of landing performance in single engine aircrafts to provide evaluations for these analyses. This can be further supported using the NASA-TLX as it would provide more support for subjective ratings of task difficulty/load compared to the SART. Lastly, our conclusions about the task performance and entropy changes seen in the difficult condition being attributed to an increase in attention and exerted effort is a fascinating finding; and one that should be further examined when cognitive resources are exceeded by task demands. In other words, both accounts for the current findings could be strengthened with additional testing to determine how task performance and gaze dispersion/sequence complexity are impacted when cognitive load is significantly higher. For instance, introducing a secondary task (i.e., dual task paradigm) to further challenge cognitive resources may reduce the extent to which additional cognitive control resources are able to modulate gaze behaviours and assist in task performance.

6.5 CONCLUSION

This work highlighted the performance, SA, and gaze behaviour differences in low-time pilots when completing a simulated landing scenario in VFR conditions with and without the presence of strong crosswinds. Traditional gaze metrics (i.e., dwell time, rate, duration, blink rate) and entropy-based metrics all provided meaningful insight about the extent to which task demands and information processing change across the different phases of flight (i.e., downwind, base, and final approach) and task difficulty. Our results suggest the changes in gaze behaviour compensated for the increased task demands and minimized the impact on task performance,

particularly during the final approach stage of landing. Lastly, the cognitive tunneling analysis remains a robust measure of task difficulty and, more importantly, pilot experience (i.e., flight hours), which accounted for up to 68% of variance in the moment-to-moment analysis of pilot monitoring behaviour (i.e., number of cognitive tunneling bouts and total bout time). In conclusion, a number of traditional and advanced gaze metrics provided critical insight into how gaze behaviour, and thus, information processing, is altered by task difficulty and task goals. However, more work is needed to validate its utility in being able to characterize pilot proficiency.

Section 7 – *Does fiducial marker visibility impact task performance and information processing in novice and low-time pilots?*

A version of this chapter has been published in *Computers and Graphics* (Ayala et al., 2024) and has been reproduced here in accordance with the Creative Commons CC-BY license.

Ayala, N., Mardanbegi, D., Zafar, A., Niechwiej-Szwedo, E., Cao, S., Kearns, S., ... & Duchowski, A. T. (2024). Does fiducial marker visibility impact task performance and information processing in novice and low-time pilots?. *Computers & Graphics*, 103889.

7.1 INTRODUCTION & BACKGROUND

Augmented Reality (AR) and Virtual Reality (VR) have gained in popularity for their ability to provide an enhanced or immersive experience that supplements or surpasses an individual's real-life environment. This is achieved through the addition of digital elements in the current environment, a complete change to a fully simulated environment, or a mixture of both these elements (Craig, 2013; Kaplan et al., 2021; Khor et al., 2016). One broad application that has exploited these technologies involves simulation-based training, which has shown great promise in increasing training efficiency while reducing the associated financial and ethical costs (Hancock, 2009). Specifically, simulator-based training environments have been used as a surrogate to learn and practice novel or dangerous scenarios that would otherwise be too risky or unethical to complete in real-life military or medical situations (McKnight et al., 2020; Rantanen et al., 2005). Other notable applications include driving, tram operation, police training and even assistive devices for people with disabilities (Caserman et al., 2022; Lee et al., 1998; Rowe et al., 2013; Velosa et al., 2020).

Two common problems experienced in AR, VR and mixed reality technologies include: (1) camera pose estimation, which is dependent on the precise localization of the camera and its projections in the environment (Duchowski et al., 2020); and (2) occlusion, when a real object appears to be occluding the virtual scene, which causes the virtual object to be rendered on the real object (Garrido-Jurado et al., 2014). The application of fiducial markers, specifically ArUco (Augmented Reality University of Cordoba) markers, has been shown to be an effective solution to these problems above other methods that rely on computer vision to identify natural features in the environment to use as reference points (Craig, 2013; Garrido-Jurado et al., 2014). However, ArUco markers are potentially distracting in their appearance within the natural environment and may interfere with natural behaviours, which is especially important with respect to the marker's usability in eye-tracking applications. To our knowledge, this paper presents the first formal investigation into how distracting visible ArUco markers are and proposes a novel invisible ArUco marker topology.

7.1.1 Paper overview and contributions

A fiducial marker system is composed of a set of artificial markers that serve as landmarks in the environment and an algorithm that performs its detection, and correction, in images (Craig, 2013). The use of fiducial markers has become increasingly popular across various domains (i.e., medicine, autonomous driving, robotics, assistive technologies, augmented reality) that rely on computer vision to complete tasks such as localization, pose estimation and gaze mapping (Craig, 2013; Duchowski et al., 2020; Garrido-Jurado et al., 2014; Hales et al., 2013; Kunz et al., 2020). The implementation of these fiducial marker systems provides the experimenter with critical pieces of information about the environment, including: (1) what computer graphic object should be presented to the user (if any) relative to marker location, (2)

what point of view should be depicted (Craig, 2013), and (3) localization of Areas Of Interest (AOIs) in the scene (relative to the marker) (Duchowski et al., 2020). Notably, the utility of this information provides answers to questions like where in the camera frame, and thus 3D world, does a given object exist, and what is its orientation and position with respect to the camera. In the context of eye-tracking applications, the answers to these crucial questions gathered through fiducial markers and their associated software libraries has improved the eye-tracking capabilities to accurately and efficiently extract gaze position in 3D space (Duchowski et al., 2020; Garrido-Jurado et al., 2014). This has become a critical area of research since advancements in eye-tracking technologies have produced affordable, high-quality mobile eye-trackers that allow researchers to study gaze behaviour in natural, real-world environments.

Eye-tracking literature that focuses on gaze behaviour is primarily concerned with where an individual is looking within the task environment, when and for how long. As such, the associated gaze analysis involves defining AOIs within the task environment to infer the deployment of attention (Ayala et al., 2022). Even though fiducial markers help identify these AOIs across static and dynamic environments (Duchowski et al., 2020; Garrido-Jurado et al., 2014; Bykowski et al., 2018), their black-and-white appearance makes for an unnatural and potentially distracting element in the environment that may interfere with the naturally occurring visual scan patterns.

This paper describes the implementation of a new physical topology of ArUco markers that enable the detection of pre-defined AOIs in 3D space via *invisible* markers using an Infra-Red (IR) light source and IR- sensitive camera. As such, we examined two different ArUco markers and evaluated their impact on gaze behaviour and pilot performance during a simulated

flight landing task. Moreover, we examined the distinct influence these markers had across the aforementioned parameters in novice and low-time pilots, respectively.

The main contributions of this paper are to: (1) develop invisible ArUco markers by utilizing an IR light source and IR-sensitive camera that can detect the marker through the longpass visible-spectrum- blocking film, and (2) provide a formal examination of the utility of invisible markers and how they may provide a more natural, imperceptible (or at least non-distracting) substitute for the potentially distracting original black and white ArUco markers in eye-tracking research.

7.1.2 Related work

Several types of fiducial markers have been previously used (Garrido-Jurado et al., 2014; Kunz et al., 2020; Wang et al., 2016), however few have the benefit of using a novel topology that renders the markers invisible to the user while still being registered by the camera.

Koike et al. (2009) initially proposed the use of transparent markers on a liquid crystal display (LCD) tabletop system by using the polarization features of the LCD and optical films. Though these transparent markers were only suitable for tabletop applications (i.e., translator, geographic, and image filter applications), later work by Koshikawa et al. (2022) proposed a novel gaze estimation technique using similar transparent markers specifically for large LCD applications. Here the transparent markers were arranged on the screen as novel reference points for eye-tracking by using illumination properties to estimate the center of the corneal sphere. It is important to note that the use of polarization-based transparent markers on LCD screens is heavily influenced by illumination conditions and that the applications are restricted to the 2D space on the LCD screen itself.

Ellsworth et al. (2016) proposed the use of a retro-reflective surface in which embedded tracking fiducial information was encoded by spatial patterns that were detected with a head mounted projected display (HMPD) system. The addition of view lenses and optics in the HMPD provided a means to create an augmented reality experience through which the user perceives the retro-reflected images and may also interact with objects in the real world. The benefit of this additional equipment provided an environment where the user's experience was dictated by the point of view (POV) generated by the HMPD. When the system detected a change in the fiducial image as a result of head movement, the system was able to render the appropriate view of the computer generated image (CGI) object from the new POV. This system is dependent on an electrical system for the Infra-Red Light Emitting Diodes (IR LEDs), which serve as their invisible reference points. As such, marker installation is more cumbersome, and its placement remains rigid in the border of the screen.

Ehambram et al. (2019) showed how marker detection could be markedly improved through the use of retro-reflective materials, especially in Infra-Red (IR) image data. They examined two different retro-reflective materials to create a novel marker topology based on ArUco markers that were detectable in RGB and IR image data. By involving an IR-sensitive camera system, the negative impact of external disturbances (i.e., changes in light conditions) on marker detection were diminished. However, it became evident that these markers needed to be placed in a well-light environment for best detection performance.

None of the proposed marker systems have directly examined the user impact of employing these invisible markers. Notably, ArUco markers serve as the most accessible and feasible fiducial system implementation option in a simulator environment as the marker and software libraries are publicly available online (Munoz-Salinas et al., 2013). Moreover, the

flexible marker placement enables accurate gaze registration and gaze mapping in 3D space (Duchowski et al., 2020).

Gaze registration is especially important with respect to eye-tracking applications involving gaze mapping in environments that extend beyond an LCD screen or tabletop with changing illumination conditions. Below, we propose the use of invisible markers and compare their influence on gaze behaviour to that of the original visible ArUco markers in novice and low-time pilots within a simulated aircraft environment.

The examination of the potential differences in the performance and gaze profiles of these two classes of pilots comes as a result of earlier work in various domains reporting that experienced performers or experts use vision differently than less experienced individuals; and thus, exhibit distinct gaze profiles (Brams et al., 2018, Gegenfurtner et al., 2011; Peißl et al., 2018; Ziv et al., 2016). Indeed, the gaze profiles of experienced performers suggest a more efficient use of vision to comprehend and integrate relevant visual information in order to produce an optimal action or sequence of actions associated with a specific task. More specifically, experienced performers tend to be better at selectively attending to task-relevant information and ignoring task-irrelevant information (Blaslev et al., 2012; Brams et al., 2018; Gegenfurtner et al., 2011; Haider et al., 1999; Spitz et al., 2016), encoding and retrieving information from memory more rapidly (i.e., extract more meaningful information from shorter fixations) (Brams et al., 2018; Gegenfurtner et al., 2011; Donovan et al., 2013, Godwin et al., 2015), and have an extended perceptual span (i.e., better at using their peripheral vision to globally analyze a scene) (Brams et al., 2018; Gegenfurtner et al., 2011; Jaarsma et al., 2015; Kundel et al., 2007; Wood et al., 2013). Therefore, the level of experience an individual has with a given task or environment is an important factor to consider. As such, we hypothesized that if

visible ArUco markers are distracting and draw attention away from the simulated landing task we expect that performance will be negatively impacted and that visual attention will reflect an increase in exploratory (i.e., task-irrelevant) behaviour.

An increase in exploratory gaze behaviour would be directly proportional to Stationary Gaze Entropy (SGE; i.e., fixation dispersion), and to Gaze Transition Entropy (GTE; i.e., visual scan sequence complexity, e.g., see Krejtz et al. [2015]). It is also expected that the subjective distraction scores and the number of fixations towards these marker AOIs will be higher in the visible marker condition compared to the invisible marker condition.

Alternatively, the presence of distracting stimuli (i.e., fiducial marker) may only impact individuals new to the environment or novice performers who are not as proficient in using their vision efficiently, and thus have a tendency to equally attend to task-relevant and task-irrelevant information. In this case, we should see that performance and gaze behaviour of low-time pilots do not change across marker conditions while the corresponding metrics in novice pilots are modulated by the visible marker condition. It should be noted that since SGE and GTE are time-averaged gaze behaviour metrics, a novel dynamic entropy analysis (Ayala et al., 2023b) will be explored here to further elucidate any gaze behaviour differences that exist.



Figure 7.1. Illustration of the ArUco marker placement within the flight simulator for both visible (left) and invisible (right) marker types.

7.2 INVISIBLE MARKERS

This paper builds on Garrido-Jurado et al.'s (2014) ArUco marker system to propose a novel invisible ArUco marker. This fiducial marker system uses the original algorithm to produce markers that maximize the inter-marker distance and the number of bit transitions. The current work uses the 4×4 marker library to achieve this (Munoz-Salinas et al., 2013). The benefit of using these square-based fiducial markers lies in the presence of four prominent points that can be employed to obtain pose, while the inner region is used for marker identification. To make the markers invisible to the viewer, a longpass filter that blocks any visible light is used to cover the marker making it look dark to the human eye but still visible to an IR-sensitive camera.

The visible markers were printed off on a piece of paper and reinforced from the back with plastic that has small magnets placed in the four corners of the marker (**Figure 7.2 (left)**). These magnets were strong enough to attach to the instrument panel as well as the four magnets embedded in the four corners of the same shape cover that can go on top of the printed marker. The front cover is an infrared long- pass filter (from Edmund Optics) that blocks any light with

wavelength outside the 700–1600 nm range and transmits more than 90% of the infrared light within that range. The markers were illuminated by a 2000 mAh rechargeable IR light source. (**Figure 7.2 (right)**) attached to the ceiling above the pilot. The use of multiple markers has been shown to increase pose estimation reliability and aid in reducing the occlusion errors (Garrido-Jurado et al., 2014; Ehambram et al., 2019). Therefore, the current work uses four markers within the cockpit environment that are attached to the instrument panel via magnets that allow for easy marker attachment and removal. Three ArUco markers (4×4 cm each) were attached to the top row of the instrument panel and one bigger marker of size 7.5 cm was attached to the empty portion of the far-right end of the panel. The arrangement of the markers across both conditions is shown in **Figure 7.1** with AOIs labeled in **Figure 7.4**.

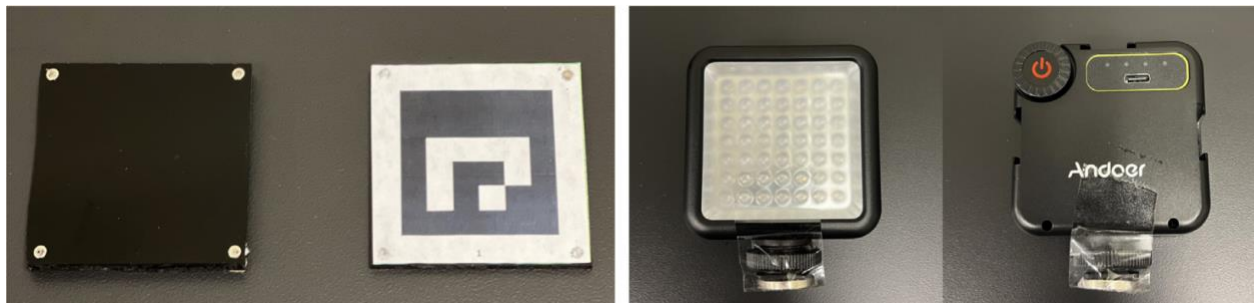


Figure 7.2. IR longpass filter marker covering (left) and IR light source used in the experiment (right). The front side of IR light source was directed toward the markers inside the simulator.

7.3 EMPIRICAL VALIDATION

We conducted experiments to evaluate the effectiveness of the invisible ArUco markers compared to visible ArUco markers in a flight simulator (AL250, ALSIM, France) with novice and low-time pilots. This study was approved by the University of Waterloo Research Ethics Board Committee (application number: 43238).

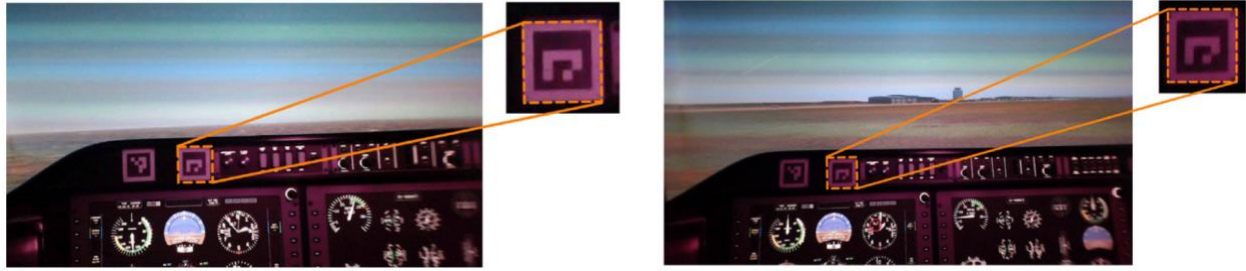


Figure 7.3. Examples of the camera view during the experiment: (left) image of the cockpit with original ArUco markers with a close-up of a marker, and (right) image of an invisible ArUco marker with close-up of the marker. Notice that the markers are visible to the camera in both instances (in right image due to IR illumination) although invisible markers appear slightly dimmer.

7.3.1 Experimental procedure

Participants first completed a visual screening including a visual acuity test using the Bailey–Lovie chart and a stereoacuity test using the Randot Stereo test (Stereo Optical Company, Inc.). Prior to commencing the experimental trials, a pilot briefing (completed by an instructor pilot) and practice trial was performed to familiarize the participants with the simulator environment and flight path (AL250, ALSIM, France). This was completed using invisible markers to ensure the cockpit looked as similar as possible to the unmodified AL250 ALSIM simulator environment that participants typically use during training. The experimental landing scenarios were programmed in the flight simulator environment, configured as a Cessna 172 (included steam-gauge instruments) flying into the Waterloo International Airport (Breslau, ON, Canada). Participants were asked to complete a total of 4 customized landing trials while their eye movements were recorded. The marker types were counterbalanced across the landing scenarios in an ABAB or BABA manner (A: visible markers; B: invisible markers). The landing

scenarios involved visual flight rules (VFR) with high visibility and low wind conditions and represented an easy ‘error-free’ landing scenario. All participants received the exact same environmental configurations. **Figure 7.4** shows a screen capture of the simulated scenario. Each trial was pre-set to start as a downwind-to-base-to-final approach to the airport at an altitude of 2017 ft, 1 nautical mile away from the downwind runway threshold with flaps and trim set to zero, and at a starting speed and RPM of approximately 110 kts and 2000 rpm, respectively.

Each participant performed a 5-point eye-tracking calibration that was completed by the examiner. This was followed by a 9-point validation that all participants passed (1.24° mean average gaze error) prior to commencing the experimental trials. At the start of every trial, each participant manually initiated the landing scenario once a 1-point eye-tracking calibration was completed by the examiner. The task goal was to land the plane as smoothly and accurately as possible relative to the center of the 500 ft markers near the start of the runway. The trial was terminated after the participant brought the plane to a complete stop, or if the landing was deemed unsuccessful (i.e., plane crash or plane landed off the runway). The participant was then asked if they had noticed and were distracted by the markers present in the cockpit environment.



Figure 7.4. Illustration of the visual stimuli employed in the simulator environment. The participant's point of view in the cockpit was pre-set for a downwind-to-base-to-final approach to Waterloo International airport, Breslau, ON, Canada. The orange boxes represent the fourteen Areas Of Interest used in the gaze analyses. These include the (1) airspeed, (2) attitude, (3) altitude, (4) VHF Omni Range (VOR), (5) heading, (6) vertical speed, (7) power, as well as (8) the front window, (9) left window, (10) right window. The additional 4 Areas of Interest (11-14) represent the four markers that were placed in the cockpit around the instrument panel for optimal camera detection.

7.3.2 Participants

We recruited 20 healthy research participants (male: 15, female: 5) with an average of age 23 years (min: 19, max: 29). The low-time pilot group consisted of 10 participants who had obtained at least their private pilot's license (PPL) (8 males, 2 females; age range: 20–25 years,

mean: 20.6 years old, SD: 1.9 years) with an average of 193 flight hours (range: 65–280 h). The novice group consisted of 10 participants who had little-to-no flight experience (7 males, 3 females; age range: 19–29 years, mean = 23.3 years old, SD = 4.4 years) with an average of 4 flight hours (range: 0–28 h). All novice pilots had knowledge about the various information given by the instruments in the cockpit (altimeter, altitude, vertical speed, etc.), and were familiar with how to manually interact with the aircraft. All participants were either current aviation students or had graduated from the aviation program at the University of Waterloo, Canada. All participants had normal or corrected-to-normal vision and had not been previously diagnosed with a neuropsychiatric/neurological disorder or learning disability. Based on previous work by Ayala et al. (2023b), a sample size of at least 16 was deemed sufficient to reach an effect size of 0.5 for the ANOVA's involving the gaze metrics being assessed in the current study, and task difficulty.

7.3.3 Experimental set-up and apparatus

Flight simulator. We used an AL250 ALSIM flight simulator (ALSIM, France) configured as a Cessna 172, see **Figure 7.1**. Like a Cessna 172, flight instruments included an instrument panel (steam gauge configuration), an avionics/GPS system, an audio/lights panel, a breaker panel, and a Flight Control Unit (FCU). Participants sat in a height-adjustable seat (left pilot seat) with their aviation headset plugged in for ATC (air traffic control) call-outs. The field of view covered by the simulator was $250^{\circ} \times 49^{\circ}$ via panoramic VFR-VR-HD projectors. The participants controlled the aircraft with a yoke, throttle lever, and rudder pedals. Stimuli presentation and behavioural data collection and acquisition were controlled from the Instructor Station and Engineering pack (ALSIM, France).

Eye tracker. A pair of AdHawk MindLink eye-tracking glasses (AdHawk Microsystems Inc., Waterloo, ON, Canada), were used to track the participants' gaze. AdHawk Mindlink is a head-mounted non-camera-based eye tracker that uses a pair of ultra compact microelectromechanical systems (MEMS) to track the eye and gaze movements (**Figure 7.5.**). The front-facing camera (82° field of view, 1080p, 30 Hz) of the Mindlink glasses were modified by the manufacturer for this study with the IR blocking filter removed from the lens thus the camera was able to detect both visible and invisible ArUco marker types (**Figure 7.3.**). The eye tracker was operating at 250 Hz, transmitting the gaze data and the video of its front-facing camera via the AdHawk eye tracking software to a laptop (60 Hz refresh rate, 1920 × 1080 pixels, Microsoft 11) visible only to the experimenter. The eye tracker was also equipped with an IMU sensor which was used to track the subjects' head rotations, which was critical for detecting when the subject was looking at the left and right windows.



Figure 7.5. AdHawk Mindlink eye tracker used in the study.

7.3.4 Data processing and analysis

Eye tracking data. In addition to the standard 3D gaze vectors that AdHawk eye tracking software exports, the ArUco markers were also detected and tracked in real-time by a software that was provided by Adhawk and the position of the markers with respect to the camera were

stored separately. The gaze data were also projected into the 2D plane of the instrument panel (gaze-in-screen) where the AOIs were defined. The gaze-in-screen data was used to determine the fixation positions with respect to the AOIs. Notably, the 3D gaze vectors were not affected by the noise from the marker detection, thus we used that for saccade and fixation detection using a post-processing script. The saccade detection algorithm was based on the algorithm proposed originally by Nyström and Holmqvist (2010) with some modifications to work on our data captured at 250 Hz. The gaze vectors were converted to gaze angles (azimuth and elevation) relative to the head. The velocity and acceleration signals were computed for each channel using a Savitzky–Golay filter with a 9 sample window size. The square root of the sum of squares of each of the two velocity channels was taken as the gaze speed used for detecting the saccades. Unlike the original method that uses a fixed saccade-peak-velocity threshold for detecting the saccade candidates, we used a low-pass filtered version of the velocity signal (rolling average with a 50 sample window) to increase the threshold in the noisy regions of the data. After classifying the eye movements into saccades and fixations, the average of the gaze-in-screen samples during each fixation was taken as the fixation position. Each fixation was then assigned to one of the AOIs if it was intersecting the AOI region in the instrument plane. The 14 AOIs chosen for the current experiment were determined a priori based on the assumption of importance and relevance to the task combined with prior knowledge of previous work in the same domain. For instance, the 7 gauges in the cockpit provide information to the pilot about the state of the aircraft. Given that the pilots were flying in VFR conditions, it was also crucial that they look outside (i.e., front, left and right windows) for any incoming traffic or other hazardous items/events. Furthermore, since the current experiment was interested in examining how the visibility of ArUco markers impacted gaze, it was determined that the last 4 AOIs should

surround the markers themselves to see if marker fixation occurred more in one condition versus the other. Notably, even though several other displays and levers were located within the cockpit and were not designated as an AOI, most do not provide the most pertinent information required to fly an aircraft. Indeed, previous work by our group has demonstrated that non-AOI dwell time typically accounts for less than 5% of total dwell time across all participants (Ayala et al., 2023b). All AOIs that were defined were relative to the markers in the panel coordinate system. It is important to note that although marker placement is more flexible with the use of ArUco markers, there were a few limitations with the current simulator environment that impacted the placement of markers. Theoretically it is better to place as many markers as possible on the instrument panel to get more information in the image that can be used for gaze estimation. However, in the simulator environment there were limited amounts of free space on the panel to place markers of sufficient size for detection. It was important for the markers to be in areas that were in front of the pilot as they would be more likely to be visible throughout flight, which is required for proper gaze estimation. The software that was used for tracking the markers did not have the coplanarity constraint, but the markers were placed on the panel and thus were coplanar. The placement of two markers on the right side of the panel covered the right side of the simulator and helped determine when pilots turned their head to the right as the front panel markers would go out of camera view. Accordingly, IMU data and markers helped identify right head turns toward the right window AOI. Since no other markers could be placed in an area that remained in frame during left head turns, left window AOI detection primarily relied on IMUs and the loss of all markers. Trial-by-trial video play back was also used to confirm that IMU data correctly identified instances where pilots looked at the left or right windows. If any data were erroneously recorded as looking outside the left/right window, it was manually adjusted.

Performance and gaze analysis. We evaluated the landing performance, gaze behaviour, and subjective level of marker distraction across the two fiducial marker types. This was also compared across novice and low-time pilots.

Landing performance was broken down into three domains: completion time (sec; duration of time from the start of the landing scenario to the plane coming to a complete stop on the runway), landing accuracy (degrees; the difference between the center of the plane and the center of the 500 ft runway marker at point of touchdown), and landing hardness (feet per minute, or fpm; the rate of descent at point of touchdown).

Gaze behaviour was examined using Stationary Gaze Entropy (SGE), Gaze Transition Entropy (GTE), a novel dynamic entropy-based analysis (i.e., gaze tunneling bout analysis) (Ayala et al., 2023b), and fixation frequency specifically on the ArUco markers. The time-averaged entropy-based analysis (i.e., SGE and GTE) was completed using 14 AOIs (**Figure 7.4.**) that were automatically discretized during pre-processing. Eye fixations in the fourteen AOIs were assigned a number from 1 to 14 indicating the AOI where the eyes fixated. The gaze tunneling bout analysis was completed during post-processing based on the methodology described by Ayala et al. (2023a). To characterize the deployment of attention between the internal cockpit environment and external scenery, all cockpit AOIs (1- 7, 11-14) were collapsed into a single “inside” AOI and all external AOIs (8-10) were collapsed into a single “outside” AOI. Custom scripts were written in Python to compute SGE and GTE, as well as complete the gaze tunneling bout analysis (Ayala et al., 2022; Ayala et al., 2023b).

SGE was computed by producing a vector, $\mathbf{V}$, of length 14, where $\mathbf{V}_i$ was the total number of fixations at AOI i . $\mathbf{V}$ was then divided by the total number of fixations in the

sequence, so that $\mathbf{V}_i$ was the probability of a fixation landing at AOI i . The probability vector $\mathbf{V}$ was then applied to the following Shannon, 1948):

$$H_{SGE}(V) = - \sum_{v \in V} v \cdot \log(v)$$

Equation 1

GTE was computed by creating a 14×14 transition matrix, $\mathbf{M}$, where $\mathbf{M}_{i,j}$ was the total number of transitions from AOI i to AOI j . Each row, $\mathbf{M}_{i,\cdot}$, was divided by the sum of row i , so that $\mathbf{M}_{i,\cdot}$ represented the probability of fixation transition from AOI i to any of the fourteen AOIs. Finally, GTE was computed using (2) (Ciuperca and Girardin, 2007), applying the transition matrix $\mathbf{M}$ and the probability vector $\mathbf{V}$ as follows:

$$H_{GTE}(M) = - \sum_{i=1}^{14} V_i \sum_{j=1}^{14} M_{i,j} \cdot \log(M_{i,j})$$

Equation 2

The gaze tunneling bout analysis was conducted by first performing a probability calculation, P(inside), using a 10 s sliding average window of assigned binary numbers. These were based on fixation AOIs that specified whether the fixation was inside (1) or outside (0) the cockpit. When P(inside) was equal to 1, participants were continuously fixating inside the cockpit for at least 10 s. When P(inside) was equal to 0, participants were continuously fixating outside the cockpit for at least 10 s. The specific ‘bout’ analysis characterized the instances when pilots stopped cycling their attention toward both internal and external environments (i.e., gaze tunneling). A gaze tunneling bout was defined as a period of time during which fixations remained entirely outside of the cockpit for at least 10 s. These bouts were detected as connected

components of zeros in the probability time series. Number of bouts was defined as the number of instances a bout was detected within a trial. Bout duration (s) was defined as the average duration of all bouts that occurred in a trial. Total bout time (s) was defined as the sum of all the individual bout durations within a trial.

Fixation frequency was computed by counting the number of fixations that landed on any of the four ArUco markers during the duration of the landing trials. Last, subjective level of marker distraction was computed by taking the average of participants responses to the question “Did you notice the markers while flying the plane and were they distracting?” (i.e., Yes=1, No=0).

7.4 RESULTS

Comparison between pilot groups (between-group factor: novice, low-time) and marker types (within-group factor: visible, invisible) on performance, gaze, and subjective measures were analyzed with a mixed model ANOVA with an alpha level set to 0.05. It should be noted that an exploratory post-facto analysis of gaze inside and outside defined AOIs confirmed that non-AOIs (i.e., secondary displays, levers in the middle console, etc.) were not looked at often by participants in either group. Similar to previous work (Ayala et al., 2023b), non-AOI dwell time accounted for less than 5% of all dwell time across all participants.

The results depicted in **Figure 7.6.** show similar performance scores across marker types and pilot groups, respectively, for completion time (sec) (novice: mean = 174.9 s, SD = 0.3; low-time: mean = 161.9 s, SD = 1.3), and landing accuracy (°) (novice: mean = 0.12°, SD = 0.005; low-time: mean = 0.053°, SD = 0.0003). Landing hardness (fpm) demonstrated a main effect of group, $F(1, 18) = 13.218$, $p = 0.002$, $\eta_p^2 = 0.423$, $d = 0.930$. Specifically, novice pilots landed

harder (mean: -221 fpm, SD: 24) than low-time pilots (mean: -115 fpm, SD: 8). No main effect or interaction involving marker type was present.

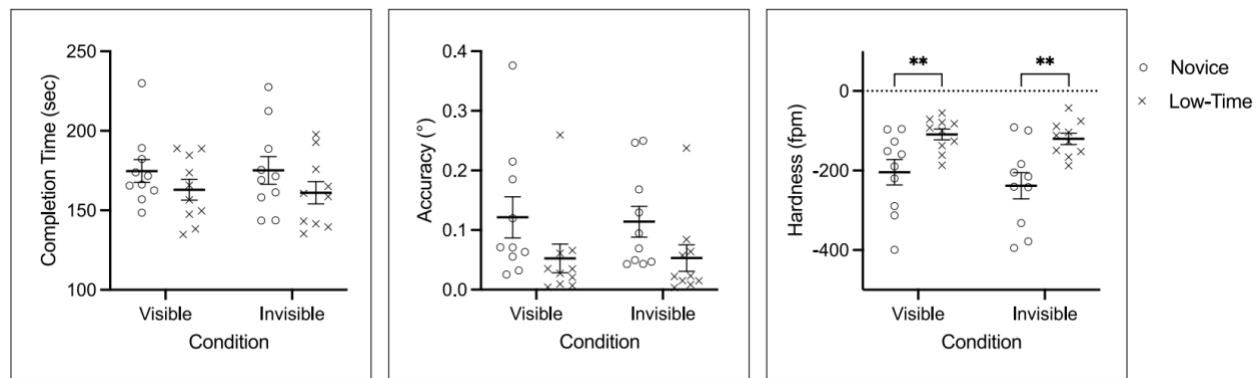


Figure 7.6. Individual participant data points and their respective group means for (left) completion time expressed in seconds, (center) landing accuracy expressed in degrees, and (right) landing hardness expressed in feet per minute (fpm). Error bars represent SEM (* $p < 0.05$, ** $p < 0.01$).

SGE (bits) demonstrated a main effect of pilot group with gaze dispersion being more dispersed amongst low-time pilots (mean: 2.79 bits, SD: 0.01) compared to novice pilots (mean: 2.29 bits, SD: 0.08), $F(1, 18) = 13.855$, $p = 0.002$, $\eta_p^2 = 0.435$, $d' = 0.940$ (**Figure 7.7. (left)**). GTE (bits) also demonstrated a main effect of pilot group with visual scan sequences being significantly more complex amongst low-time pilots (mean: 1.92 bits, SD: 0.01) than novice pilots (mean: 1.38 bits, SD: 0.06), $F(1, 18) = 13.161$, $p = 0.002$, $\eta_p^2 = 0.422$, $d' = 0.929$ (**Figure 7.7. (center)**).

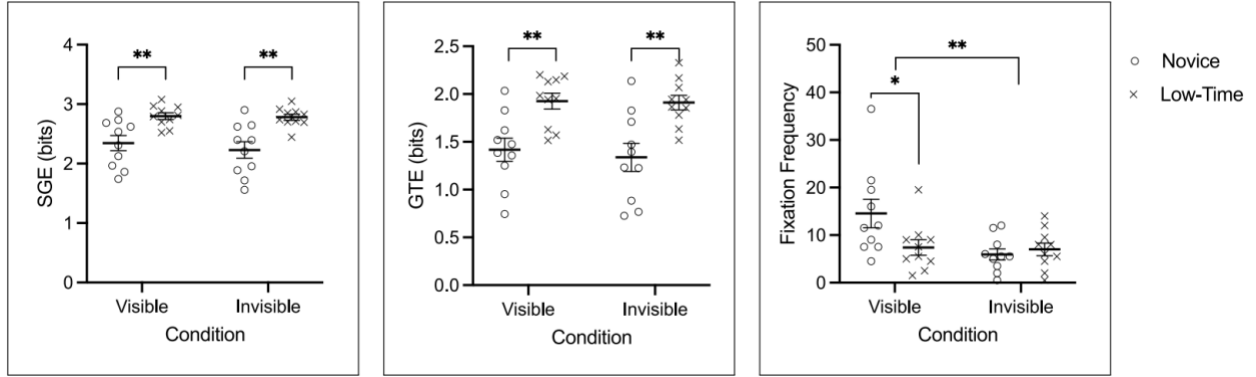


Figure 7.7. Individual participant data points and their respective group means for (left) Stationary Gaze Entropy (SGE), expressed in bits, (center) Gaze Transition Entropy (GTE), expressed in bits, and (right) fixation frequency (i.e., count). Error bars represent SEM (* $p < 0.05$, ** $p < 0.01$).

Number of bouts demonstrated a main effect of pilot group with gaze tunneling bouts occurring more frequently in novice pilots (mean: 2.2, SD: 1.2) than low-time pilots (mean: 1.0, SD: 0.4), $F(1,18) = 9.893$, $p = 0.006$, $\eta_p^2 = 0.355$, $d = 0.845$ (**Figure 7.8. (left)**). Average bout duration (sec) was not significantly impacted by pilot group, $F(1, 18) = 2.043$, $p = 0.170$, $\eta_p^2 = 0.102$, $d = 0.273$, or marker type $F(1, 18) = 0.099$, $p = 0.757$, $\eta_p^2 = 0.005$, $d = 0.060$. Total bout time (sec) demonstrated a main effect of pilot group with total gaze tunneling bout time being significantly longer in novice pilots (mean: 45.03 s, SD: 27.63) than low-time pilots (mean: 17.15 s, SD: 8.87), $F(1, 18) = 9.944$, $p = 0.005$, $\eta_p^2 = 0.356$, $d = 0.846$ (**Figure 7.8. (right)**).

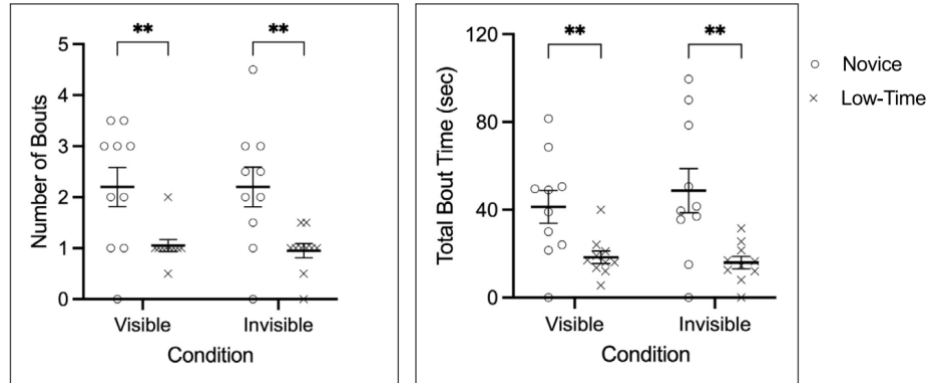


Figure 7.8. Individual participant data points and their respective group means for (left) number of gaze tunneling bouts (i.e., count), and (right) total gaze tunneling bout time (sec). Error bars represent SEM (* $p < 0.05$, ** $p < 0.01$).

Fixation frequency demonstrated a main effect of marker type with fixations occurring more frequently at marker AOIs in the visible marker condition (mean: 11, SD: 8) compared to the invisible marker condition (mean: 6, SD: 4), $F(1, 18) = 8.563$, $p = 0.009$, $\eta_p^2 = 0.322$, $d' = 0.790$. Notably, an interaction between pilot group and marker type was also significant, $F(1, 18) = 7.109$, $p = 0.016$, $\eta_p^2 = 0.283$, $d' = 0.713$. **Figure 7.7. (right)** specifically demonstrates the interaction where novice pilots fixated on the markers much more frequently in the visible marker condition (mean: 15, SD: 9) than the invisible marker condition (mean: 6, SD: 4). In contrast, the low-time pilots had equivocal fixations on marker AOIs across both visible (mean: 7, SD: 5) and invisible (mean: 7, SD: 4) marker conditions.

Subjective scores of marker distraction yielded a main effect of marker type, $F(1, 18) = 39.600$, $p < 0.001$, $\eta_p^2 = 0.688$, $d' = 0.999$. Both pilot groups subjectively reported that visible markers (mean: 70% SD: 41) were noticeable and/or distracting during operation of the simulator, while invisible markers (mean: 3% SD: 11) were less discernible (**Figure 7.9.**).

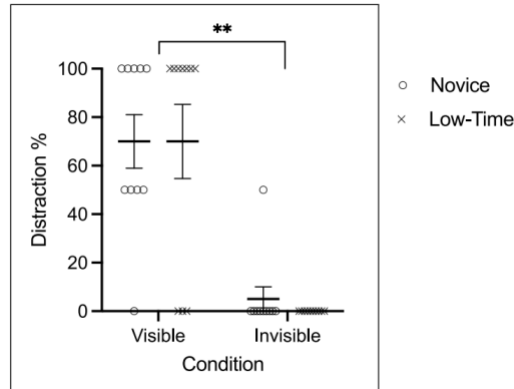


Figure 7.9. Individual participant data points and their respective group means for subjective distraction scores. Error bars represent SEM (* $p < 0.05$, ** $p < 0.01$).

7.5 DISCUSSION

The implementation and utility of novel invisible ArUco markers were confirmed by empirical results. Results showed that the marker type did not modulate any of the performance or entropy-based measures. However, the visible markers were associated with higher subjective distraction scores, and more importantly, a higher fixation frequency specifically by novice pilots.

The analysis of landing performance demonstrated that the type of ArUco marker used did not show any significant difference in task performance. However, landing performance did show differences in landing competency between the novice and low-time pilot groups. Though this did not support our hypotheses on the impact of *distracting* visible ArUco markers, it can be an indication that the landing task employed was “too simple”; a finding that suggests that the shift in participants’ attention away from the task (i.e., increase in fixation frequency toward visible ArUco markers) was not sufficient to impair landing performance during the flying conditions used here.

The analysis of gaze entropy, specifically the dispersion (SGE) and sequence (GTE) of participant scan patterns, have been shown to be effective indicators of attentional deployment (i.e., focal versus exploratory modes of attention) as well as the complexity of the scanning pattern structure (Ayala et al., 2022; Krejtz et al., 2014; Shiferaw et al., 2019). The present study demonstrated that gaze dispersion and visual scan sequence complexity were not significantly different between the type of ArUco marker used across either group of pilots. In accounting for this discrepancy, it is important to note that the marker AOIs account for a small fraction of attended areas in the environment. Indeed, over the course of a landing trial, participants spent less than 2% of performance time attending to these markers. As such, a change in the SGE or GTE—gaze metrics that examine the overall change in the probability of attending to any of the fourteen AOIs—would be unlikely to shift in response to a change that specifically revolved around the marker AOIs. Notably, the dynamic entropy-based analysis on gaze tunneling also failed to show a significant change in gaze behaviour between ArUco marker types. Though this was included as an exploratory analysis to see how gaze dynamics would be impacted, one expectation would have been to see a reduction in gaze tunneling bouts, durations or total bout time as a result of gaze tunneling bouts being interrupted by shifts in gaze toward the ‘distracting’ visible ArUco markers. Instead, the absence of such an effect suggests that the distracting gaze shifts towards the visible ArUco markers might have occurred during shifts in gaze that preceded or followed a gaze tunneling event.

The current SGE and GTE results substantiate previous work that reported well-known differences in the visual scan patterns of novice performers and expert performers (Shiferaw et al., 2019; Vickers and Williams, 2017; Ziv, 2016). Specifically, gaze dispersion (SGE) and visual scan complexity (GTE) were higher in low- time pilots compared to novice pilots.

Additionally, the gaze tunneling bout analysis found that the number of gaze tunneling bouts and total gaze tunneling time were significantly longer among novice pilots compared to their low-time pilot counterparts. In other words, novice pilots were sampling information within their cockpit environment less frequently and less effectively than low-time pilots. Together, these findings further support the argument that more experienced pilots use vision more efficiently and effectively by attending to a larger number of AOIs using a more complex scan pattern (i.e., frequently alternating between AOIs in a non-homogeneous pattern) (Brams et al., 2018; Di Nocera, 2007; Harris, et al., 1986; Peißl et al., 2018; Sarter et al., 2007; Ziv, 2016). Given that the focus of this research was on low-time pilots (i.e., less than 300 flight time hours), these results serve as a novel contribution as most research examining gaze differences in novice and expert pilots has been completed with pilots who were better classified as experienced and less experienced pilots, or expert pilots (i.e., military/commercial airline pilots) and a group of individuals with no flying experience (for review see Peißl et al. [2018] and Ziv [2016]).

Fixation frequency, particularly on the marker AOIs was also analyzed to focus the gaze analysis on the specific changes in attention deployment toward the markers themselves (i.e., task irrelevant areas). Our findings supported our alternative hypothesis, which specifically expected the visible markers to be a distraction only to novice performers. The results showed that novice pilots fixated on the markers significantly more than low-time pilots. It is important to note that this effect was driven by the fact that novice pilots fixated the markers significantly more in the visible condition compared to the invisible condition; which was more on par with the marker fixation frequency of low-time pilots. In contrast, low-time pilots attended to the marker AOIs equally across both marker conditions. These results are further supported by the subjective assessment scores of marker distraction, which showed that both novice and low-time

pilots noticed the visible markers in the environment and were, to some extent, subjectively distracted by them. Despite the low-time pilot group subjectively indicating that the visible markers were distracting, it is evident that they were better at selectively attending to the other AOIs that provided information about the operation of the aircraft itself. Additional support for such a conclusion could also be provided with evidence showing that practice with visible ArUco markers in the task environment does not impact the reported results. Although the current study was not set up to appropriately assess the effect of practice on visible ArUco marker fixation frequency due to the low trial count per condition, it is worth noting that our preliminary analysis demonstrated no consistent change in visible ArUco marker fixation frequency from the first trial to the second trial. Increasing the number of trials in future research to see if participants are able to get used to visible ArUco markers over time would be worth investigating. This could help determine if increased exposure to the visible markers through additional practice at the beginning of an experiment is enough to reduce the fixation frequency difference reported between visible and invisible ArUco markers. Furthermore, during our preliminary analysis of the effects of practice over time between trials, we also explored the effects of practice within trials. In other words, we became interested in determining if the fixation of visible ArUco markers was more frequent at the beginning of trials and if it decreased over time. Such a finding could also support the notion that exposure to the ArUco markers is required prior to experimentation to reduce the impact on gaze behaviour. Our exploratory analysis on this consistently demonstrated that the majority of visible and invisible ArUco marker fixations occurred in the later stages of the trial when the primary focus of interest was directly above and below two of the ArUco marker positions. This suggests that ArUco marker position may also play a role in the extent to which marker visibility impacts gaze behaviour. As such future work

should also look at evaluating the effect of marker placement (i.e., near areas of high task importance versus areas of low task importance) on gaze behaviour, especially in environments that offer more space and freedom for marker placement to sufficiently examine this question. We expect that if the markers are placed at locations that are more likely to interrupt visual information processing (i.e., locations closer to flight instrument gauges), the advantages of invisible markers might become more significant. Nevertheless, our results clearly indicate that low-time pilots (i.e., more experienced performers) avoided task-irrelevant areas and focused their attention more on task-relevant areas compared to novice pilots (Brams et al., 2018; Helleberg et al., 2003; Kirby et al., 2014).

7.6 LIMITATIONS AND FUTURE WORK

This study provides a novel contribution regarding the use of an invisible ArUco marker system. However, the current results are constrained by at least four limitations that should be addressed in future research. First, the landing scenario generated for this experiment was very simplistic and involved weather conditions that did not challenge the participants. As a result, the performance findings generated from the ALSIM simulator demonstrated a ceiling effect in that all participants landed the plane safely across all trials. The addition of more difficult weather conditions (i.e., strong winds, low visibility) would allow for more performance variability amongst participants and might alter the degree to which distracting stimuli in the environment (i.e., fiducial markers) impact flight performance and gaze measures. Second, it is unclear how the IR light source will need to be implemented in areas that are larger than the simulated cockpit environment used here. The detection of the invisible markers is dependent on sufficient light reaching the markers to enable camera detection. Therefore, IR-light source placement should be

investigated further to see how these markers can be used in larger open environments. Third, the subjective distraction questionnaire was neither expansive in providing a range of options for the degree of distraction, nor was it clear on whether it was asking about marker perception or marker distraction. The simplistic nature of the scale (i.e., yes or no) limited the extent to which we could infer the degree of participant distraction and also restricted the analysis that could have been completed to strengthen the conclusions regarding fixation frequency and gaze distraction. For instance, using a level of distraction scale of 0-9 instead of yes or no could have allowed for a correlation analysis to be completed on fixation frequency and flight hours. Moreover, the aspect of marker perception (i.e., the reception and interpretation of stimuli in the environment) is quite different from marker distraction (i.e., diversion of attention due to a task irrelevant object/event that consequently reduces task awareness, decision making, and performance) (Hedlund, 2006). Although participants perceived the markers it does not equate distraction. The current study did not demonstrate impairments to performance, which would suggest that visible markers are, in fact, not a distraction. However, it is important to note that some driving simulation studies have shown that performance does not always get negatively impacted by distracting stimuli, while gaze typically is impacted (i.e., increased fixation count on task irrelevant stimuli, reduced on-road gaze time, reduced gaze position variability) (Garrison and Williams, 2013). However, given the limitations of the task discussed earlier more research is required to clarify how distracting the visible markers truly are since the data did show that fixation frequency on marker locations increased when they were visible. Lastly, driving simulation studies demonstrating decrements to performance during the introduction of distracting tasks and stimuli collected continuous streams of data related to driver control (i.e., heading, speed, lane departure, steering angle, brake pressure means and variability) (Bornard et

al., 2011; Harbluk et al., 2007; Marti et al., 2021). In contrast, the current study focused on final landing metrics that were likely insufficient to capture the entirety of the pilot's behavioural performance relating to aircraft control that might have been impacted because of increased fixations on task-irrelevant stimuli (i.e., visible markers), specifically in novice pilots. A consequence of this in the current study, is that it is difficult to determine if an increase of 10–20 fixations is problematic or insignificant to the actual task at hand. Notably, some driving studies have suggested that as few as 3 fixations on task-irrelevant stimuli is sufficient to suggest that the driver became distracted (Garrison and Williams, 2013). As such, future research should be conducted with more challenging real-world tasks, continuous measures of aircraft control, and with a more complex scale for visual perception and visual distraction, respectively.

7.7 CONCLUSION

In this work, we propose a novel invisible ArUco marker topology, and also provide a formal investigation into the impact visible ArUco markers have on the behavioural performance and gaze behaviour of pilots in a simulated aircraft environment. Novice and low-time pilots completed a simulated landing scenario while their eye movements were tracked with the AdHawk Mindlink glasses. Automatic gaze mapping and AOI detection was completed using visible and invisible ArUco markers that were set up in the simulator environment. Landing performance, static and dynamic gaze entropy-based measures were not impacted by marker type. However, fixation frequency and subjective distraction scores demonstrated that visible markers were disrupting to participants and interrupted natural gaze behaviours, particularly in novice pilots. As such, our findings demonstrate that the proposed invisible ArUco marker topology serves as a more natural substitute for the original black and white ArUco marker

system, and should especially be used when there is a need for these markers to remain inconspicuous (i.e., eye-tracking applications).

Section 8 – *The effects of a dual task on gaze behaviour examined during a simulated flight in low-time pilots (from Ayala et al., [submitted]; Frontiers in Psychology)*

8.1 INTRODUCTION

Today's intelligent aircraft cockpit provides multisource information via an array of instrument panels and controls that are associated with different procedures that a pilot must attend to at any given moment (i.e., pilot monitoring) (Federal Aviation Administration, 2021). The capacity limitations associated with human information processing make the task of acquiring information from these various areas of interest (AOIs) challenging, particularly during states of high cognitive load (i.e., high amounts of cognitive resources demanded from the operator by competing activities) (Engström et al., 2017). Therefore, it is of no surprise that ineffective pilot monitoring was cited as being a major contributor to human factor errors during highly demanding phases of flight (i.e., take-off, approach and landing) (Boeing, 2021; National Transportation Safety Board, 1994). Evidently, a pilot's ability to select and monitor relevant information in such a dynamic environment are critical to flight success and safety. Therefore, being able to objectively track and quantify changes in information processing under varying cognitive load conditions is an important line of inquiry as it provides a framework to understand how operators effectively monitor task-critical sources of information necessary for successful task performance during states of high cognitive load.

Vision plays a critical role in how we interact with our environment. Eye tracking serves as a non-invasive method that reveals discrete cognitive processes and strategies used to facilitate skill performance that are not readily observable through overt behaviours alone (Ayala et al., 2022; Vickers and Williams, 2017). Indeed, where we look (i.e., fixate) is governed by

ongoing perceptual (i.e., bottom-up, lower-level stimulus features) and cognitive (i.e., top-down, higher-level mental hierarchies and goals) processes that facilitate the selection and processing of relevant information that support skill performance (Engström et al., 2017; Kendal and Schwartz, 2012; Land and Hayhoe, 2001). As such, the analysis of eye movements provides a window into what information is being processed, how efficiently that information is being processed, as well as how effectively it is integrated into pilot decision making and action generation.

Numerous studies have shown that problem solving, decision making and visuomotor behaviour of an operator are significantly impacted by changes in cognitive load (Allsop et al., 2014; Engstrom et al., 2017; Kanaan and Moacdieh, 2021; Van de Merwe et al., 2015; Yang et al., 2018). For instance, high cognitive load has been shown to be associated with increased misjudgment (i.e., spatial disorientation, increased time and error rates associated with hazard object/event detection), reduced aircraft control (i.e., increased variability in aircraft speed, pitch and turn coordinator), and unstable approaches and landings (Crosby and Parkinson, 2007; Engstrom et al., 2017; Ziv, 2016). These changes in performance were reported in parallel with a narrowing of attention toward fewer areas of interest (AOIs), more frequent fixations and longer fixation durations (Allsop and Gray, 2014; Di Nocera et al., 2007; Engstrom et al., 2017; Kanaan and Moacdieh, 2021; Lu et al., 2020; Lutnyk et al., 2023; Reimer, 2009; Wright et al., 2014). Taken together, the combined performance and gaze behaviour findings demonstrate that increases in cognitive load are associated with impairments to performance, alterations in information processing, and potentially pose a threat to flight safety.

The Cognitive Control Hypothesis states that cognitive load will selectively impair performance on tasks that require *controlled* processing (i.e., novel, uncertain, nonroutine tasks that require top-down executive functions like working memory), while leaving tasks that require

automatic processing (i.e., practiced, effortless, generally unconscious tasks that require little to no cognitive control oversight) or rather *efficient* processing relatively unchanged (Engström et al., 2017). Since the majority of pilot monitoring errors occur during cognitively demanding phases of flight (i.e., take-off, approach and landing) (Boeing, 2021; National Transportation Safety Board, 1994), the Cognitive Control Hypothesis suggests that these phases of flight are associated with controlled performance (i.e., require cognitive resources to manage the respective sub-tasks, while maintaining aircraft control) and are susceptible to increases in cognitive load, more so than any other flight phase (i.e., cruise). Accordingly, eye tracking could serve as a useful method to understand, 1) how changes in cognitive load impact information processing via changes in gaze behaviour, and 2) determining if these changes in gaze behaviour impact processing of task-relevant information critical to successful flight performance.

Notably, previous research has examined the question of cognitive load primarily by employing a dual task paradigm wherein the difficulty of performing a primary task (i.e., flying) was made more challenging by the addition of a simultaneous secondary task (i.e., visual search, mental arithmetic, auditory tone, etc.); thereby, introducing additional rules and goals to be maintained/manipulated in working memory (Engström et al., 2017). This paper aimed to build on previous aviation-related work employing a dual task paradigm to examine the utility of gaze metrics in assessing the impact of cognitive load on pilots' performance and information processing using a more ecological approach. There are two major aspects through which this study differs from previous work. First, the current work sought to use a dual task paradigm to objectively probe the impact of a secondary task on cognitive load across two phases of a flight circuit (i.e., cruise [low cognitive load], approach and landing [high cognitive load]). Since it has been well established that the two stages of flight (cruise, approach and landing) vary in the

extent to which they require cognitive resources to monitor and complete the required sub-tasks (Di Nocera et al., 2007; Federal Aviation Administration, 2001; Wilson, 2002), it is important to examine how the imposed change in cognitive load impacts the phase-specific spatiotemporal aspects of gaze behaviour. Crucially, accurately monitoring pilot's cognitive load through continuous, objective, gaze-based metrics is of great significance for the development of real-time metrics to assess pilot's information processing capability and may also be an effective method to track pilot's progression in the development of efficient information processing throughout training. Second, the current work was designed to have high fidelity to increase the generalizability of the findings to low-time pilots. Specifically, the dual task involved monitoring auditory inputs to simulate the modality of Air Traffic Control [ATC] calls, while simulating the flight environment with a fully immersive flight simulator. This design is distinct from previous work that used video recordings for the primary task (van de Merwe et al., 2012), introduced secondary tasks that did not resemble general aviation-like procedures (i.e., mental arithmetic, working memory tasks, visual search, military specific shooting tasks, etc.) (Abel, 2009; Allsop et al., 2016; Babu et al., 2019), and recruited participants with no aviation experience (Allsop and Gray, 2014; Lutnyk et al., 2023).

The current study used a secondary task as a probe to evaluate the impact of cognitive load on flight performance and gaze behaviour across two phases of flight: cruise and landing. It was hypothesized that performance on the auditory oddball will be significantly less accurate and slower during the approach and landing phase of flight compared to the cruise phase. Moreover, based on previous work using a VFR (Visual Flight Rules) paradigm, it was hypothesized that increased cognitive load will be associated with a reduction in attention allocation across a smaller set of task-relevant AOIs manifested as an increase in dwell time and fixation durations

toward the external environment (i.e., front window) (Allsop and Gray, 2014; Allsop et al., 2016; Kanaan and Moacdieh, 2021; Yang et al., 2018; Yang et al., 2020). Although more dynamic measures of gaze behaviour have only been examined in the context of increasing visuomotor task difficulty (i.e., wind manipulations) (Ayala et al., 2023b), it was expected that an increase in cognitive load would similarly be associated with a reduction in the dispersion of fixations (Stationary Gaze Entropy: SGE), a reduction in visual scan pattern complexity (Gaze Transition Entropy: GTE), and an increase in cognitive tunneling (i.e., the frequency and duration of poor internal cockpit monitoring events). These gaze behaviour changes are expected to be specifically evident during the approach and landing rather than the cruise phase due to the increased cognitive load induced by processing task-specific information to ensure safe landing, which may overlap with the processing demands of the secondary auditory task.

8.2 METHODS

8.2.1 Participants

Twenty-two participants (14 males, 8 females; age range 18-24 years, mean= 21 years old, SD= 2) were recruited from the student populations at the University of Waterloo. All participants were current aviation students or individuals who had obtained at least their private pilot's license (PPL) (number of flight hours range: 48-340 hours, mean= 199 hrs, SD= 73; PPL= 16; CPL= 6). All participants had normal or corrected-to-normal vision and had not been previously diagnosed with a neuropsychiatric/neurological disorder or learning disability. Participation was voluntary, and participants received \$25/hour as remuneration. The University of Waterloo Research Ethics Board Committee (#43564) approved the study protocol, which was performed in accordance with the 2008 Declaration of Helsinki. Consent was obtained prior to

beginning the protocol. Note that one participant was excluded from analysis as a result of corrupt data files.

8.2.2 Experimental setup and apparatus

Flight Simulator. An AL250 ALSIM flight simulator (ALSIM, France) configured as a generic single engine aircraft that is representative of a Cessna 172 was used with the necessary instrument panel (steam gauge configuration), an avionics/GPS system, an audio/lights panel, a breaker panel, and a Flight Control Unit (FCU) (see **Figure 8.1**). The field of view covered by the simulator was 250° by 49° via panoramic VFR-VR-HD projectors. The participants sat in a height-adjustable seat and controlled the aircraft with a yoke, throttle lever, and rudder pedals. Stimuli presentation, and behavioural data collection and acquisition were controlled from the Instructor Station and Engineering pack (ALSIM, France).



Figure 8.1. Illustration of the visual stimuli employed in the AL250 flight simulator environment. The participants point of view of the cockpit replicated that of a pilot flying a Cessna 172, pre-set for a cruise and landing approach to Waterloo International Airport, Breslau, Ontario, CA. The orange boxes represent the ten main areas of interest used in the gaze analyses. These include the airspeed (1), attitude (2), altimeter (3), turn coordinator (4), heading (5), vertical speed (6) and power (7) indicators, as well as the front (8), left (9), and right windows (10).

Eye Tracker. MindLink eye-tracking glasses (AdHawk Microsystems Inc., Waterloo, ON, Canada) were used to track the participants' gaze (i.e., eye and head) movements. MindLink

is a non-camera-based eye tracker embedded in a frame of eyeglasses that uses an ultra-compact micro-electromechanical system (MEMS) to track the eye and gaze movements (Zafar et al., 2023). The eye tracker was operating at 250 Hz, transmitting the gaze data and the video of its front-facing camera (82° field of view, 1080p, 30 Hz) via the AdHawk eye tracking software to a laptop (60 Hz refresh rate, 1920 x 1080 pixels, Microsoft 11) visible only to the experimenter.

8.2.3 Scenario and task

Participants were tested in a single experimental session (~90 minutes) which started with a visual screening that included a visual acuity test (Bailey-Lovie chart) and a stereoacuity test (Randot Stereo test, Stereo Optical Company Inc.). A pilot briefing (led by an instructor pilot) and practice trial were completed to familiarize the participants with the simulator environment (AL 250, ALSIM, France), flight path, and flight parameters (cruise airspeed [110 kts] and altitude [2017 ft], landing approach airspeed [65 kts] and touchdown reference point [center of 500-foot marker]). The experimental scenarios were programmed in the flight simulator environment, flying into the Region of Waterloo International Airport (CYKF; Runway 26), Breslau, Ontario, Canada. Participants were then asked to complete a total of 9 customized trials while their gaze movements were recorded. The trials were pseudo-randomized (i.e., ABC, BCA, CBA, BAC) into three types of scenarios: 1) single-task auditory oddball task (control condition), 2) single-task flying circuit (cruise [low cognitive load] and landing [high cognitive load]), and 3) dual task paradigm (single task circuit with auditory oddball task). Prior to beginning the experimental trials, a 9-point eye-tracking calibration and validation procedures were completed by the examiner (average gaze error <2°).

For the flight task, all participants received the exact same environmental configurations in Visual Flight Rules (VFR) conditions (i.e., high visibility [>20 miles] and calm [0 kts, 0°]) with a pre-set start at the beginning of the cruise stage of the flight circuit to the airport at an altitude of 2017 ft, with flaps and trim set to zero, and a starting speed and RPM of 110 kts and 2000 rpm, respectively. The flying task was segmented into two cognitive load phases. Cruise was the first phase of flight wherein the goal was to maintain aircraft control (i.e., aircraft altitude [2017 ft], airspeed [100-110 kts], and vertical speed [0 fpm] by monitoring the gauges for any slight deviations in speed or altitude and adjusting the flight controls as necessary. The second phase consisted of landing the plane wherein the goal was to land the plane as smoothly and accurately as possible relative to the center of the 500 ft markers near the start of runway 26. Note that this phase consists of several aircraft configuration changes that required specific, time-sensitive maneuvers that allowed the plane to descend and become aligned with the runway in a safe and stable manner. The trial was terminated after the participant brought the plane to a complete stop, or if the landing was deemed unsuccessful (i.e., plane crash or plane landed off the runway). Note that pilots were required to communicate with Air Traffic Control (ATC) during both phases of flight, as they normally would in a real aircraft. **Figure 8.2** illustrates the simulated scenario along with the flight path and circuit parameters.

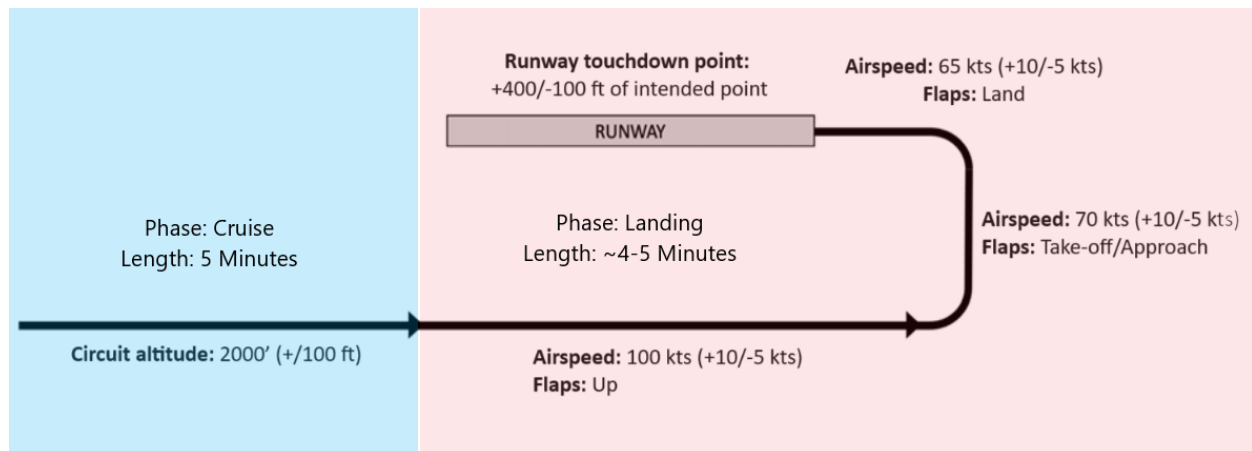


Figure 8.2. Illustration of the simulated flight path and circuit parameters for cruise (easy [blue]) and landing (difficult[pink]) phases of flight.

Auditory oddball task (control task) used auditory tones that were created using VPIxx 3.2.1 software and were presented biaurally at 80 decibels using Apple computer speakers. Participants remained in the left pilot seat of the cockpit with the aircraft parked on the runway. Before testing, participants were presented with an iteration of high (1000 Hz) and low (375 Hz) tones and asked if they could discriminate between the two tones. Participants were then instructed to respond by pressing on the pilot push-to-talk button located on the yoke with their left hand as quickly as possible when hearing a high “target” tone (probability= 30%), and to withhold a response when hearing a low “non-target” tone (probability = 70%). Participants were notified when a trial began by the instructor followed by an iteration of high and low tones randomly presented at inter-stimulus intervals ranging from 1900 msec to 2100 msec.

The dual task paradigm condition involved performing both the flying circuit and auditory oddball tasks simultaneously. Participants were informed by the instructor when the

trial began. The auditory tones were presented and responded to just as they were during the auditory only condition. Every participant was given the same instructions to “primarily focus on flying the aircraft in the dual task condition; therefore, if the task becomes too demanding you should focus on flying the aircraft more than attending to the auditory task.”

After each experimental trial (i.e., single flying circuit, single auditory oddball task, dual task paradigm) the participant was then asked to complete the Situational Awareness Rating Technique (SART) questionnaire to gauge their subjective opinion on various domains related to task difficulty, and the supply and demand of attentional resources required during task performance (Taylor, 1990).

8.2.4 Data processing and analysis

Situation awareness (SA) was assessed using a subjective questionnaire. The SART questionnaire (Taylor, 1990) is a post-trial self-report questionnaire that uses a 7-point Likert scale (1= Low; 7= High) across 10 dimensions of SA. This is collapsed into three larger dimensions of attentional demands, attentional supply and situation understanding. These ratings are then combined to calculate an overall measure of SA.

$$SA = Understanding - (Demand - Supply)$$

Equation 1

SART scores were post-trial surveys that were completed at the end of a trial and thus could not be partitioned into cruise and landing flight phases. As such, SART scores were subjected to a repeated measures ANOVA (within-group factor: task condition [single, dual]).

Based on the various flight performance measures collected, analysis was split across the two phases of flight: 1) Cruise performance- includes airspeed and vertical speed mean and variability- and 2) Landing performance- includes completion time (sec; duration of time from the start of the flying scenario to the plane coming to a complete stop on the runway), landing accuracy (degrees; the difference between the center of the plane and the center of the 500 ft runway marker at point of touchdown). Notably, the distinct cruise and landing performance metrics were not continuous across phases. For instance, vertical speed would only be an appropriate measure for landing performance during the final phase of landing but not necessarily all other parts of the landing phase (i.e., downwind, base). Additionally, the required mean airspeed was expected to constantly decrease throughout the landing phase whereas during cruise, there is an expected consistent mean airspeed that participants are expected to maintain. All flight metrics were examined within their respective phase of flight and subjected to a repeated measures ANOVA with a single within-subject factor (task condition: single, dual).

Auditory oddball task performance was assessed using auditory response time (msec; response time to accurate target tones) and accuracy % (percent; percentage of correct responses to total number of tones presented). Auditory oddball performance was examined during single auditory task conditions, as well as the dual task condition. As a result of the dual task condition was split into a cruise (low load) and landing (high load) phase of flight, raw auditory response data were subjected to a repeated measures ANOVA (within-group factor: cognitive load [single auditory task, dual cruise, dual landing]). To specifically examine the impact the dual task had on auditory responses during the different flight phases, the dual task 'cost' (DTC) for auditory response metrics were calculated by expressing participants' changes in the respective measures

during dual task performance relative to their control task (i.e., auditory only) performance according to the formula (Zhou et al., 2014; Manor et al., 2010; Schwenk et al., 2010):

$$DTC = \left[\frac{(\text{dual task performance} - \text{control task performance})}{\text{control task performance}} \right] \times 100$$

Equation 2

Dual task cost comparisons for auditory response were conducted with a repeated measures ANOVA (within-group factor: phase of flight [cruise, landing]).

Gaze data were post-processed offline using a custom-made script that used the 3D gaze vectors provided by AdHawk software for saccade and fixation detection, similar to Ayala et al (2024). Eye-movement traces were visualized by the experimenter and played back at a slowed speed superimposed over the video displaying the simulator environment. The task environment was discretized using a custom code that partitioned the simulator environment into ten areas of interest (AOIs) (**Figure 8.1**). The AOIs were manually defined to represent airspeed (1), attitude (2), altimeter (3), turn coordinator (4), heading (5), vertical speed (6), and power (7) cockpit indicators, along with the front (8), left (9), and right (10) windows. In line with previous work, fixations found outside these AOIs were defined as non-areas of interest and were excluded from analysis (<5%). Trials with missing data (i.e., loss of signal >30% found in <8% of trials) and outliers from each of the dependent variables (i.e., >1.5 the interquartile range around the first and third quartile found in <1% of trials) were removed.

Gaze behaviour was examined across each phase of flight (cruise [low load], landing [high load]) using traditional and advanced (i.e., static and dynamic) gaze measures (Ayala et al., 2022; Ayala et al., 2023b; Ayala et al., 2024; Glaholt, 2014). Traditional gaze-based analysis included the calculation of dwell time (%; total dwell duration spent within a given AOI as a

function of total flight time) and average dwell duration (msec; average duration of uninterrupted dwells within a given AOI). Static entropy-based analyses were completed using the ten AOIs outlined in **Figure 8.1** (Ayala et al., 2023b). Custom scripts (Ayala et al., 2023b) were used to compute both SGE (Equation 3) (Shannon, 1948) and GTE (Equation 4) (Ciuperca and Girardin, 2007), which were then normalized (Equation 5) (Shannon, 1948).

Specifically, SGE was computed by first producing a vector, $\mathbf{V}$, of length 10 (for each AOI), where $\mathbf{V}_i$ was the total number of fixations at AOI i . $\mathbf{V}$ was then divided by the total number of fixations in the sequence, so that $\mathbf{V}_i$ was the probability of a fixation landing at AOI i . The probability vector $\mathbf{V}$ was then applied to Equation 3.

$$H_{SGE}(V) = - \sum_{v \in V} v \cdot \log(v)$$

Equation 3

GTE was computed by first creating a 10x10 (AOI) transition matrix, $\mathbf{M}$, where $\mathbf{M}_{i,j}$ was the total number of transitions from AOI i to AOI j . Each row, $\mathbf{M}_{i,*}$, was divided by the sum of row i , so that $\mathbf{M}_{i,*}$ represented the probability of fixation transition from AOI i to any of the ten AOIs. Finally, GTE was computed using Equation 4 (Ciuperca and Girardin, 2007), applying the transition matrix $\mathbf{M}$ and the probability vector $\mathbf{V}$.

$$H_{GTE}(M) = - \sum_{i=1}^{10} V_i \sum_{j=1}^{10} M_{i,j} \cdot \log(M_{i,j})$$

Equation 4

$$H_{NORMAL} = H/H_{MAX}$$

Equation 5

The dynamic entropy-based analysis was completed using the 10 second average sliding window developed by Ayala et al (2023). This cognitive tunneling analysis was used to examine the reduction in gaze transitions from the external environment to the internal cockpit environment. This analysis specifically quantified the number of cognitive tunneling bouts (i.e., number of instances in which a dwell remained entirely outside of the cockpit for at least 10 seconds), bout duration (sec; the average duration of all bouts that occurred in a trial), and total bout time (sec; the sum of all individual cognitive tunneling bout durations within a trial). Gaze behaviour metrics were collected during both phases of flight (cruise, landing) and under single flight task and dual task conditions. Therefore, these data were subjected to a 2x2 repeated measures ANOVA (within-subject factors: Task Condition [single, dual], Phase of Flight [cruise, landing]).

All ANOVAs were performed with an alpha level set at 0.05. The Bonferroni post hoc correction for multiple comparisons was also applied for all post hoc analyses following the repeated measure ANOVAs to determine significant differences between variables. Table 8.1 provides a summary of all the ANOVAs conducted and their respective independent variables.

Table 1. List of dependent measures and their associated ANOVAs and independent variables		
Dependent Variable	ANOVA	Independent Variables
Flight Performance Metrics	1x2	Task Condition (single flying, dual task)
Raw Auditory Performance Metrics	1x3	Cognitive Load (single auditory, dual cruise, dual landing)
Dual-Task Cost Auditory Scores	1x2	Phase of Flight (cruise, landing)
Gaze Behaviour Metrics	2x2	Task Condition (single flying, dual task)
		Phase of Flight (cruise, landing)
SART Scores	1x2	Task Condition (single flying, dual task)

Note: The listed dependent variables represent the larger clusters of actual dependent variables. Flight performance metrics include landing performance metrics (total completion time, landing error, landing hardness), as well as cruise performance metrics (airspeed and vertical speed mean and variability). Raw auditory performance metrics include response time and response accuracy. Dual task cost auditory scores include auditory response cost and auditory accuracy cost. Gaze behaviour metrics include dwell time, fixation duration, SGE, GTE, cognitive tunneling bout number, duration, and total time. SART scores include overall SA, SA supply, SA demand, and SA understanding scores.

8.3 RESULTS

8.3.1 Subjective Situation Awareness Ratings

SART scores were examined with a repeated measures ANOVA with a single within-subject factor (task condition: single, dual). Participant self-ratings of situation awareness demonstrated a main effect of task, $F(1,20)=31.199$, $p<0.001$, $\eta_p^2=0.609$. Participants reported higher subjective situation awareness in the single flying condition (mean= 21.5, SD= 5.0) than in the dual task (mean= 18.0, SD= 4.6) (**Figure 8.3**). The breakdown of situation awareness scoring also demonstrated main effects of task for attentional demand and attentional supply, $F(1,20)=49.021$ and 7.827 , $p<0.001$ and 0.012 , $\eta_p^2=0.710$ and 0.281 , respectively (**Figure 8.3**). Specifically, attentional demand demonstrated a significant increase from the single flying condition (mean= 6.30, SD= 3.17) to the dual task (mean= 10.73, SD= 2.95). This increase in attentional demand was seen in parallel with a significant increase in attentional supply from the

single (mean= 17.69, SD= 4.03) to the dual task (mean= 19.28, SD= 2.69). No significant differences were reported for task understanding ($p=0.054$).

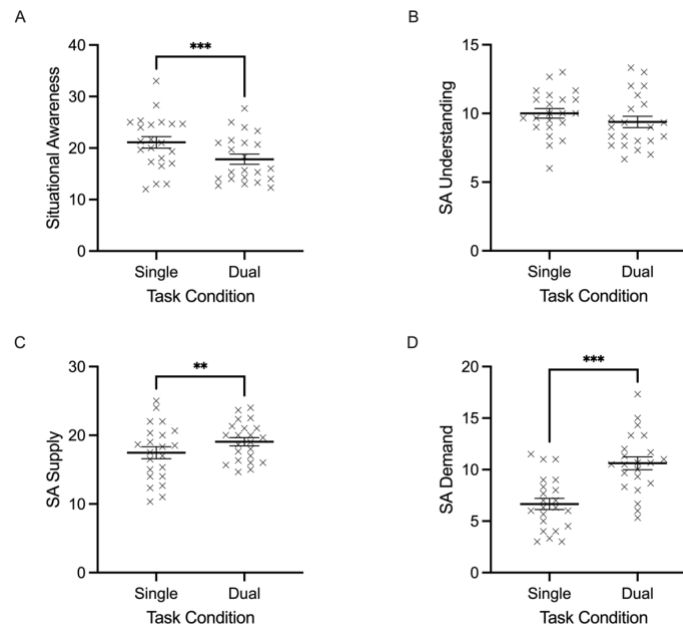


Figure 8.3. Individual data points and their respective group means for situation awareness (SA) (A), SA understanding (B), SA supply (C), and SA demand are demonstrated for single flying and dual task conditions. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

8.3.2 Auditory task performance

Raw auditory performance scores were subjected to a repeated measures ANOVA with a single within-subject factor (cognitive load: single, dual cruise, dual landing). Dual task cost scores were subjected to a repeated measures ANOVA with phase of flight (cruise, landing) as the only within-subject factor.

Auditory oddball task performance results demonstrated a significant main effect of cognitive load for response time and response accuracy, $F(2, 40) = 30.914$ and 23.648 , $p < 0.001$

and <0.001 , $\eta_p^2=0.607$ and 0.542 , respectively. Specifically, response time increased as a function of cognitive load (single= 752.17 msec, SD= 393.16; dual cruise= 1079.26 msec, SD= 487.44; dual landing= 1450.71 msec, SD= 574.79). Response accuracy decreased as a function of cognitive load (single=99.6%, SD= 0.8; dual cruise= 98.3%, SD= 2.9, dual landing= 90.9%, SD= 7.5).

Figure 8.4AB illustrate the dual task costs associated with auditory response time and response accuracy, which demonstrated a main effect of phase of flight, $F(1,20)= 6.106$ and 23.039 , $p=0.027$ and <0.001 , $\eta_p^2=0.234$ and 0.535 , respectively. Specifically, response time cost increased significantly from cruise (mean= 71.2%, SD= 102.7) to landing (mean= 166.3%, SD= 205.3). This finding is consistent with the response accuracy cost (i.e., cost is reflected as the inverse of the actual accuracy difference to express increased ‘cost’ to performance), which is also shown to increase significantly from cruise (mean= 1.3%, SD= 3.1) to landing (mean= 8.7%, SD= 7.6).

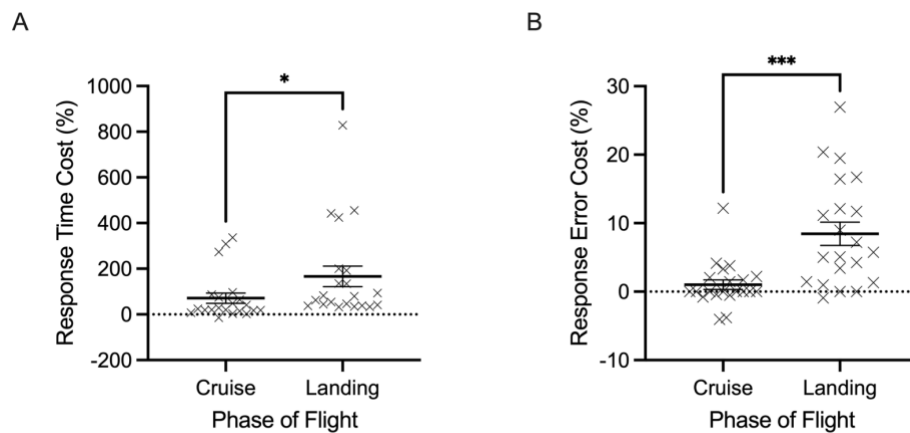


Figure 8.4. Individual data points and their respective group means for auditory response time cost (%) (A) and response error cost (%) are demonstrated for cruise and landing phases of flight. Error bars represent SEM. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

8.3.3 Flight task performance

Flight performance measures were subjected to a repeated measures ANOVA (within-group factor: task condition [single, dual task]). Cruise aircraft performance results demonstrated no significant differences between the single and dual task for cruise airspeed mean and variability, $F(1, 20)= 1.594$ and 0.091 , $p= 0.224$ and 0.767 , $\eta_p^2=0.087$ and 0.005 , or cruise vertical speed mean and variability, $F(1, 20)=0.613$ and 1.407 , $p=0.444$ and 0.252 , $\eta_p^2=0.035$ and 0.076 , respectively.

Landing performance results demonstrated no significant differences between the single and dual task for completion time, $F(1, 20)= 1.308$, $p= 0.266$, $\eta_p^2= 0.061$, and landing accuracy, $F(1, 20)= 0.126$, $p=0.727$, $\eta_p^2=0.006$. However, landing hardness was significantly higher (i.e., more negative) in the dual task, $F(1, 20)= 8.562$, $p=0.008$, $\eta_p^2=0.300$. All means and standard deviations for cruise and landing flight performance measures are reported in Table 8.2.

Table 2. Flight performance values for single and dual task conditions		
Dependent Variable	Single	Dual
Completion time (sec)	494.1 (22.2)	491.2 (25.35)
Landing accuracy (°)	0.055 (0.059)	0.054 (0.049)
Landing hardness (fpm)	-95.56 (46.3)**	-117.1 (51.35)**
Cruise airspeed (kts)	104.4 (0.85)	104.6 (0.83)
Cruise airspeed variability	0.50 (0.44)	0.46 (0.37)
Cruise vertical speed (fpm)	-3.77 (15.53)	-3.14 (14.64)
Cruise vertical speed variability	2.21 (1.90)	2.92 (2.16)

*Note: Mean (standard deviation) flight performance values are reported for single and dual task conditions. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.*

8.3.4. Gaze Behaviour

Gaze metrics were assessed via a 2x2 repeated measures ANOVA (within-group factors: task condition [single, dual task], phase of flight [cruise, landing]). Dwell time (%) demonstrated a main effect of flight phase across several AOIs (Table 8.3). Data reveal significant increases in dwell time towards airspeed and left window AOIs, as well as significant reductions in dwell time across altimeter, heading, and front window AOIs from the cruise phase to the landing phase of flight. All dwell time means, standard deviations (SD), and statistical test values for the main effect of phase of flight are presented in Table 8.3. Notably, dwell time toward the power gauge demonstrated a main effect of task, $F(1,20)=7.417$, $p=0.013$, $\eta_p^2=0.271$. Specifically, dwell time was higher during the dual task (mean= 6.67%, SD= 0.73) compared to the single task (mean= 5.41%, SD= 0.49). Dwell time toward the turn coordinator demonstrated an interaction between task and flight phase, $F(1,20)= 6.689$, $p=0.018$, $\eta_p^2=0.251$. Post-hoc decomposition of this interaction demonstrated that dwell time toward the turn coordinator was significantly higher for the cruise phase during the flying only (i.e., single) task (mean = 8.58%, SD= 4.34) compared to the dual task (mean= 6.87%, SD= 2.86), while turn coordinator dwell time remained relatively consistent across tasks during the landing flight phase (single: mean= 6.61%, SD= 3.47; dual: mean= 6.29%, SD= 3.01). No other effects of flight phase, task condition, or interactions between the two reached statistical significance ($p>0.05$). The lack of main effect or interaction involving task condition for dwell time are illustrated in **Figure 8.5A**. As such the mean (SD) and group statistics were collapsed across task conditions and reported in Table 8.3 for each phase of flight.

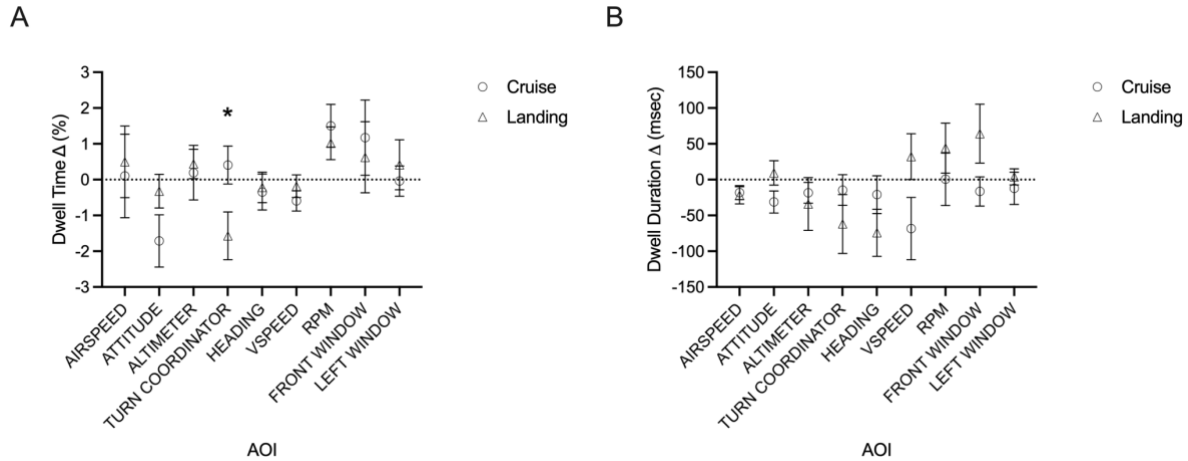


Figure 8.5. Group means for dwell time difference (%) (A) and dwell duration (msec) difference (B) (Δ difference: dual-single) across all areas of interest (AOIs) during cruise (open circle) and landing (open triangle) phases of flight. Error bars represent SEM. Significant interactions between task and flight phase are indicated accordingly * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table 3. Dwell Time (%) Descriptive and Group Statistics					
Area of Interest (AOI)	Cruise	Landing	F Statistic	P Value	Eta Squared
Airspeed	6.77 (5.45)	16.43 (7.06)	77.293	<0.001	0.794
Altimeter	12.72 (6.14)	6.12 (2.47)	44.94	<0.001	0.692
Heading	4.99 (4.39)	2.52 (1.48)	9.547	0.006	0.323
Front Window	40.88 (15.31)	35.37 (7.47)	6.241	0.021	0.238
Left Window	4.61 (2.38)	16.42 (5.59)	119.21	<0.001	0.856

Note: Mean (standard deviation) dwell time (%) values across significant areas of interest (AOI) are reported across cruise and landing phases of flight, which were collapsed across single and dual task conditions since no main effect of task condition was reported. Statistical test values (F , p , η_p^2) for main effects of phase of flight (i.e., cruise, landing) are reported for only significant AOIs. Df (1,20).

Results for average dwell duration are shown in Table 8.4. There was a main effect of flight phase across several AOIs. Specifically, significant increases in fixation duration across airspeed, turn coordinator, and front window AOIs, along with a significant decrease in fixation duration toward the altimeter AOI. Airspeed also demonstrated a main effect of task, $F(1,20)=5.983$, $p=0.024$, $\eta_p^2=0.024$, which specifically demonstrated a reduction in fixation duration during the dual task (mean= 480.72 msec, SD= 73.42) compared to the flying only condition (mean= 500.67 msec, SD= 87.39). All other AOIs were not significantly modulated by task condition or phase of flight ($p>0.05$). The lack of main effect or interaction involving task condition for dwell duration are illustrated in **Figure 8.5B**. As such the mean (SD) and group statistics were collapsed across task conditions and reported in Table 8.4 for each phase of flight.

Table 4. Average dwell duration (msec) Descriptive and Group Statistics (SEM)					
Area of Interest (AOI)	Cruise	Landing	F Statistic	P Value	Eta Squared
Airspeed	428.88 (85.73)	552.51 (107.18)	24.657	<0.001	0.552
Altimeter	502.67 (101.32)	457.99 (104.67)	4.985	0.037	0.2
Turn Coordinator	431.69 (146.78)	496.93 (146.77)	5.88	0.025	0.227
Front Window	657.22 (172.90)	888.88 (284.48)	12.572	0.002	0.386

Note: Mean (standard deviation) average dwell duration (msec) values across significant areas of interest (AOI) are reported across cruise and landing phases of flight. Statistical test values (F , p , η_p^2) for main effects of phase of flight (i.e., cruise, landing) are reported for only significant AOIs. $Df(1,20)$.

SGE (bits) showed a main effect of flight phase, $F(1,20)= 5.311$, $p=0.0332$, $\eta_p^2=0.210$, demonstrating that fixation dispersion was significantly greater during landing (mean= 2.74 bits, SD= 0.16) compared to the cruise phase (mean= 2.66 bits, SD= 0.24) (**Figure 8.6A**). However, SGE and GTE were not significantly impacted by task, and there was no interaction with flight phase ($p>0.05$). Similarly, there were no significant main effects of task, flight phase or interaction for cognitive tunneling bout metrics ($p>0.05$) (**Figure 8.6**).

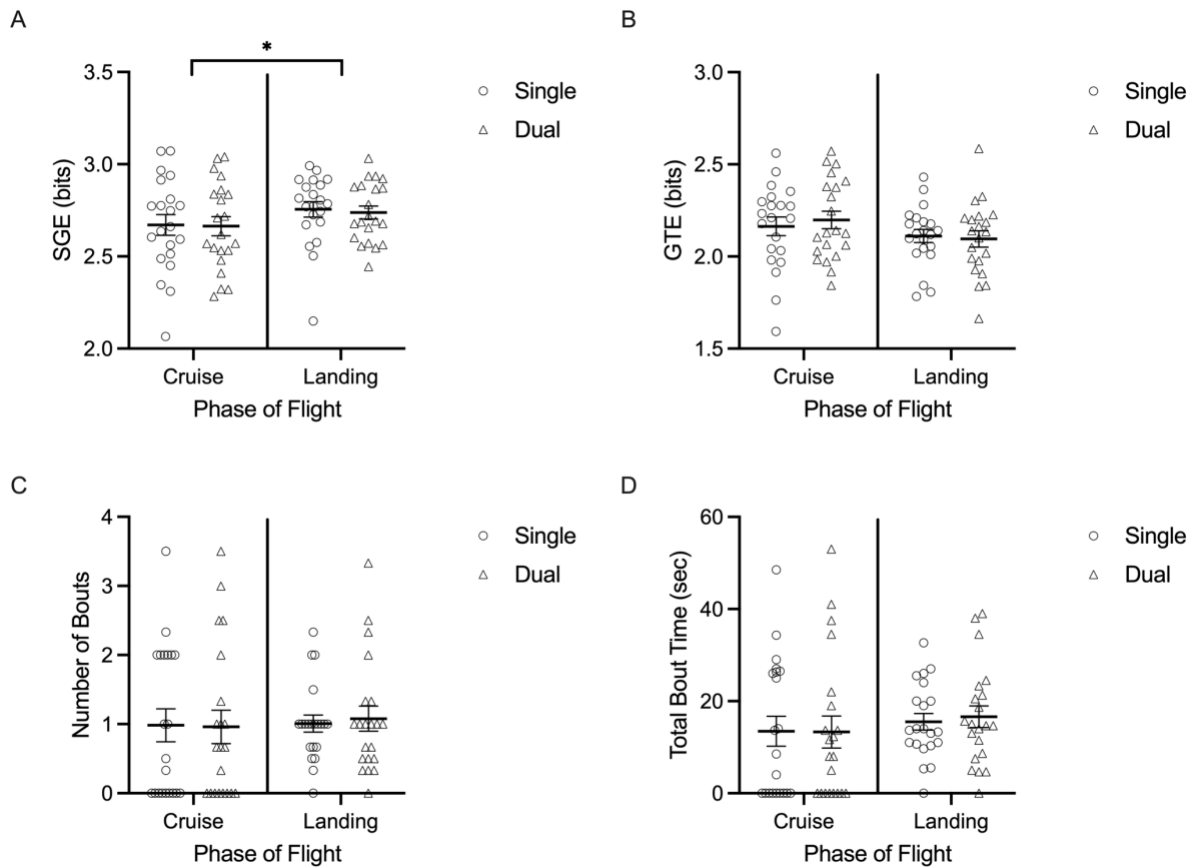


Figure 8.6. Individual data points and their respective group means for stationary gaze entropy (SGE) (A), gaze transition entropy (GTE) (B), number of cognitive tunneling bouts (C), and total cognitive tunneling bout time (sec) (D) are demonstrated for both task conditions (single, dual)

across cruise and landing phases of flight. Error bars represent SEM. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

8.4 DISCUSSION

Pilot monitoring behaviour is a critical aspect of safe aircraft operation. Understanding where and when pilots allocate their attention are important aspects of information processing. Specifically, gaze analysis could help to understand a pilot's ability to use task-relevant information in the environment to assess, plan and generate the necessary sequence of actions required at any given moment. The current study examined the effect of a dual task on flight performance and gaze behaviour during two phases of flight with differing levels of cognitive load. Participants reported the dual task was more challenging as indicated by the SART questionnaire. Notably, the auditory task revealed that the landing phase imposed a higher cognitive load compared to the cruise phase as demonstrated by reduced accuracy and longer response times to auditory tones. In contrast to our hypotheses, flight performance and most of the gaze behaviour indicators were not significantly affected by the dual task suggesting that the pilot cohort tested in this study attained a level of proficiency that allowed them to efficiently deploy their cognitive resources to complete the flying task. A parallel lack of difference between the tasks for gaze metrics indicates efficient information processing in the dual task condition. These findings are discussed in the context of evaluating gaze behaviour as a marker of information processing efficiency.

Although SA and auditory task performance confirmed the dual task increased cognitive load as hypothesized, responding to the auditory tones had very small effects on flight performance and gaze behaviour metrics in both the cruise and landing flight phases. The flight

performance results confirm that participants followed instructions to make flying the aircraft the primary task but indicate that the secondary task was not challenging enough to impair performance on the primary task, as has been demonstrated in other literature employing the dual-task paradigm (Allsop and Gray, 2014; Babu et al., 2020; Engstrom et al., 2017; Lutnyk et al., 2023; Yang et al., 2018; Yang et al., 2020). The only significant difference in flight performance was an increase in landing hardness during the dual task. However, the approximately 22 fpm increase in landing hardness is not perceptible, nor did it deem the landings unsafe given that the aircraft in question can handle landings with a landing hardness of 700 fpm.

With respect to the gaze behaviour findings, few significant differences were reported between single and dual tasks. First, average dwell durations were shown to decrease significantly (~20 msec) during the dual task, specifically for the airspeed gauge. According to previous work, this finding suggests that participants required less time to process airspeed AOI information in this condition (Ayala et al., 2022; Brams et al., 2018; Glaholt, 2014). However, this reduction in dwell time was only shown across a single AOI, which does not provide sufficient evidence that information processing became more efficient in general during the dual task condition. Second, dwell time was significantly higher toward the power gauge during the dual task condition. This increase in power dwell time was reported alongside an interaction effect for the turn coordinator gauge. Specifically, it was demonstrated that dwell time was highest for the turn coordinator during the single task cruise flight phase. Therefore, it seems as though the dual task condition caused a shifting of attention from the turn coordinator to the power gauge. This shift in gaze, however, is only shown to be true during cruise, which is an interesting finding considering that the turn coordinator is primarily used as a visual reference to

gauge the rate of turn (or lack of turn during cruise) whereas the power gauge is primarily used to get a sense of engine power and fuel efficiency (Federal Aviation Administration, 2001). As a result of no significant reductions in dwell time shown across any other AOIs because of task condition, it is likely that it came from a number of AOIs, which collectively resulted in an increase in dwell time towards the power gauge during approach and landing. In any case, since these changes in dwell time were quite small (~1-1.5%) and they were not accompanied by a significant reduction in fixation distribution (SGE), there is no evidence to suggest that these changes were associated with a narrowing of attention toward fewer task critical AOIs during states of high cognitive load. Taken together, the reported changes in gaze behaviour may reflect minor changes in information processing that helped maintain flight performance and scanning of task critical AOIs across single and dual task conditions during cruise. However, the significance of these changes in gaze behaviour and their impact on task performance is not clear and require further research.

In contrast to the traditional gaze metrics, the more dynamic measures of gaze dispersion (SGE), gaze sequence complexity (GTE), and cognitive tunneling (i.e., poor gaze monitoring characterized by a lack of instrument scanning and long dwells outside the cockpit) demonstrated no significant differences between single and dual tasks, nor did they show any interaction with phase of flight to suggest that the task condition had a significantly larger impact on gaze during the approach and landing phase of flight. Instead, the distribution of fixations (SGE) demonstrated an increase in the allocation of attention across a wider range of AOIs during approach and landing compared to cruise, which was associated with a number of changes in traditional measures of gaze behaviour (i.e., dwell time and fixation duration). However, given that these changes were not associated with the increase in cognitive load from the task condition

manipulation or an interaction with phase of flight, we propose that the findings reflect the changes in task objectives and sub-goals between cruise and landing. These findings are in stark contrast to previous work employing a similar set of gaze metrics, which found that increases in task difficulty (i.e., significant crosswinds) (Ayala et al., 2023b; Ayala et al., 2024) and cognitive load (i.e., dual task paradigms) (Engström et al., 2017; Kanaan and Moacdieh, 2021; Lu et al., 2020; Yang et al., 2018; Yang et al., 2020) resulted in a narrowing of attention during high difficulty/cognitive load task conditions via reductions in SGE, GTE, saccade amplitudes, visual scanning speed (scan path length per second), along with increases in the percentage of time looking at the center of the external environment (i.e., center of the road or the front window) and cognitive tunneling behaviour. A key distinguishing feature between previous dual task work and the current work is the type of secondary task employed. Unlike previous studies that used tasks that are not typically practiced or encountered often in traditional domain-specific routines (i.e., mental arithmetic tasks or the n-back task) (Crosby and Parkinson, 2007; Engtsröm et al., 2017; Kanaan and Moacdiej, 2021; Lu et al., 2020; Yang et al., 2018; Yang et al., 2020), the current auditory task closely resembled the aspect of responding to aircraft specific ATC calls while ignoring ATC calls to other aircraft. It is important to note that the actual ATC calls and verbal responses were replaced with target and non-target tones that required a simple button press. Accordingly, the lack of verbal responses and memory of relevant call signs (i.e., Whiskey-Alpha-Tango-52) might have resulted in the addition of a secondary task without necessarily increasing the demands on working memory that might have been necessary to see performance deficits in the primary task (Bhojwani et al., 2022; Engström et al., 2017; Lavie, 2005). Therefore, we propose that the cognitive load manipulation was not sufficiently

challenging to warrant any significant deviations from pilot scan patterns between the cognitive load manipulations.

In accounting for this discrepancy, the Cognitive Control Hypothesis (Engström et al., 2017) provides support for an alternative explanation for the lack of performance and gaze findings reported in the current work. Namely, the apparent inconsistencies demonstrated may be reconciled if we were to consider the flight scenario employed to be a task associated with more efficient (i.e., automatic) performance and not controlled performance. As a result of the recruitment pool consisting of pilots who all obtained at least their PPL - thus implying a certain level of proficiency with flying a single engine aircraft-, perhaps their experience with flight circuits (including cruise and landing phases of flight, which are normally accompanied by numerous auditory-verbal ATC calls) may be characterized as well-practiced maneuvers that are refined over the course of training. Thus, it is not a stretch to suggest that task performance is supported during challenging scenarios by improved information processing efficiency among low-time pilots who obtain their PPL/CPL certifications. Since the Cognitive Control Hypothesis states that manipulations to cognitive load selectively impairs controlled performance, the current study likely demonstrates a case wherein increasing cognitive load through a secondary task left the automated flying subtasks and pilot gaze behaviour mostly unaffected. Indeed, work by Nibbeling and colleagues (2012) demonstrated similar findings when examining the effects of anxiety and a dual task paradigm on a dart throwing task across novice and expert players. Specifically, primary task performance, perceived effort and gaze behaviour were negatively impacted by anxiety and cognitive load in novice players whereas expert players experienced little-to-no changes in performance and perceived effort. Although, expert players also demonstrated negative gaze behaviour changes, it was less detrimental to information processing

compared to novice players. This finding was taken to evince that expert players developed a level of information processing and performance efficiency that spared the use of cognitive resources, which were then used to support task performance when cognitive load and anxiety increased. Notably, research examining the extent to which improved efficiency in experts impacts gaze behaviour during states of high cognitive load is sparse and requires further investigation.

8.4.1 Limitations and future directions

Although the results suggest that the way in which the dual task paradigm was employed was not sufficiently challenging for the recruited pilot group, it provides additional questions for future research about its utility in providing an objective measure of pilot monitoring behaviour and flight proficiency. Specifically, experienced pilots who have reached a level of information processing and performance efficiency with the evaluated flight maneuvers will have spare cognitive control resources to allocate towards secondary cognitively demanding tasks. On the contrary, less experienced pilots who are still performing the same flight maneuvers at a controlled level of performance will have fewer cognitive control resources to allocate toward a secondary cognitively demanding task and will likely show larger task-related and information processing impairments. The extent to which such an impairment will become evident in pilot monitoring behaviours via gaze analyses requires further exploration. Yet, this work provides a necessary empirical stepping stone from which auditory ATC call outs can be used in place of auditory tones to not only increase cognitive load sufficiently, but also to increase fidelity (i.e.,

mimicking a busy airspace with frequent ATC calls being conducted to several aircraft). Furthermore, a more significant limitation associated with the current work is connected to the way in which auditory response time data were collected. The auditory tones were created and recorded using an external device VPixx, while the hardware used to collect the responses was implemented in the ALSIM simulator. Thus, the synchronization of the auditory tones with responses was not perfectly synchronized. Future work should incorporate stimulus presentation and response recording in the simulator environment to get a more accurate recording of auditory tone responses. Despite this limitation, it is important to note that the auditory response accuracy data reported in the current study echoes previous work in demonstrating the presence a dual task cost that is further exaggerated under more demanding task conditions. This is a finding that was not contaminated by the synchronization limitation and thus provides support in the auditory task findings reported here (Allsop and Gray, 2014; Allsop et al., 2016; Crosby and Parkinson, 2007; Lutnyk et al., 2023; Yang et al., 2020).

8.5 CONCLUSION

The present paper investigated the utility of gaze behaviour in assessing changes in cognitive load within a simulated flying task using a dual task paradigm. Auditory task performance demonstrated that increases in cognitive load differentially impacted the cruise and landing phases of flight. Specifically, dual task cost in response time and response errors was higher in the approach and landing phase of flight compared to cruise. The increase in task demands across flight phases was associated with an increase in average dwell duration across AOIs that were crucial to the phase specific sub-tasks. However, since most gaze behaviour metrics and flight performance remained unchanged when the dual task was introduced it is

likely that the recruited pilot group was not sufficiently challenged. These findings provide guidance for future research in this area with regard to tasks that are not effective in creating measurable levels of cognitive load at the low-time pilot level, while also shedding light on the efficacy of utilizing dual task paradigms to assess pilot information processing and flight proficiency.

Section 9 – *Summary of Research*

9.1 GENERAL DISCUSSION

The research presented in this thesis explored the utility of gaze behaviour as an indicator of information processing effectiveness during complex skill performance in ecological environments. Results from the five studies provide three novel contributions, which are discussed in the context of the information processing and skill performance framework depicted in **Figure 3.1**. First, the current work thoroughly investigated and evaluated an extensive set of gaze metrics while manipulating task demands, including sensorimotor difficulty and cognitive load, an individual's resource availability (i.e., expertise), and environmental contexts. Although some of the metrics have been independently examined previously, this document outlines the first set of experiments that comprehensively examines the spatial and temporal aspects of gaze behaviour during the performance of complex tasks in ecological contexts. Second, the examination of error trials (i.e., non-optimal task completion, crashes) in studies one and two provided valuable insight into aspects of information processing that may have negatively impacted task performance. This work was critical in demonstrating how gaze behaviour could be used to identify how and when participants deviated from optimal behaviours, which has implications for training and reducing the risk of human factor related errors. Third, the latter four studies incorporated a novel gaze behaviour metric (i.e., cognitive tunneling) that sought to capture the more dynamic aspects of scanning behaviour related to ignoring, forgetting, or not having the capacity to look at a specific set of task-critical AOIs. These contributions are further discussed below while highlighting different use-cases for how gaze can be effectively employed to gain insight into information processing across complex ecological applications.

First, the findings reported here are in line with previous work showing the utility of eye movements as a window to examine information processing mechanisms related to skill performance that behavioural measures alone could not reveal. In using the information processing and skill performance framework to test this assertion in more naturalistic environments, the core component of all studies presented here was the comprehensive assessment of several gaze metrics based on changes to the task demands, environmental contexts, and an individual's resource availability. Results revealed that the use of traditional (i.e., dwell time, fixation rate, fixation duration, scan path length) and dynamic (i.e., SGE, GTE) gaze behaviour metrics provided important temporal and spatial information related to the underlying cognitive processes supporting complex task performance in neurocognitive and aviation related assessments. For instance, study one demonstrated that SGE was higher in erroneous trials than in successful trials, indicating a lack of top-down selective attention processes that should have prioritized gaze towards task-relevant AOIs instead of task-irrelevant AOIs— particularly during the initial planning stage of problem solving. These findings also provided evidence for the utility of additional dynamic measures of gaze behaviour (i.e., SGE and GTE) as important indicators of changes in fixation distribution and scanning complexity. Collectively, these measures demonstrated how the focus of attention became more biased toward areas that held more task-critical information and were necessary for successful task performance; a finding that became more evident as task requirements (i.e., sensorimotor demands, cognitive load) increased. Thus, providing evidence that gaze behaviour is an effective measure of changes to information processing resulting from manipulations of task demands. The robustness of the current findings is supported by changes in gaze behaviour that were consistently seen across two different environmental contexts (i.e., low-fidelity and high-fidelity

cockpit environments). Furthermore, this work also revealed the extent to which an individual's resource capacity, defined by expertise, impacts gaze behaviour. Specifically, our findings demonstrated that despite notable ceiling effects in performance, ab-initio and more experienced (low-time) pilots differed significantly across several gaze behaviour measures. Specifically, experienced pilots used vision more efficiently and effectively; a finding that was made apparent through the more dispersed gaze patterns that were much more complex and had fewer instances of cognitive tunneling and fixations directed toward task-irrelevant stimuli compared to ab-initio participants. As such, I believe the work conducted here provides strong support for examining gaze behaviour as an objective and effective method to evaluate and monitor changes in information processing during the performance of complex tasks in naturalistic environments.

The second major contribution of this work stems from the analysis of gaze behaviour during unsuccessful task performance. Indeed, current practice for training and performance assessment in high-risk environments including medicine, aviation, military, and even day-to-day driving environments relies on the subjective assessment of human behaviour. While expert evaluators may be able to effectively determine how well an individual can perform a given task, gaze behaviour stands to provide critical insight related to when and how information processing fails. For example, study one and two revealed that unsuccessful trials were associated with a lack of selective attention being allocated to task-critical AOIs. Although the application of these measures to real-world environments requires further research, the current work paves the path towards the applications of using such gaze metrics as a tool for training in simulation environments. In particular, the recording and analysis of gaze behaviour may help instructors identify issues in visual scanning that can be trained and improved upon to enhance skill

performance. Notably, the actual impact of such training on reducing human error requires research.

The third major contribution of the current work is the development of an objective and quantifiable measure of cognitive tunnelling. The construct of cognitive tunneling has been described in the literature as a type of gaze behaviour reflective of poor instrument scanning, which was hypothesized to be linked to poorer flight performance (Allsop and Gray, 2014; Allsop et al., 2017; Xion et al., 2016). However, previous studies did not provide an empirical quantification of cognitive tunneling, which is the major contribution of this thesis. Specifically, the experimental results provide strong support for the effectiveness of the cognitive tunneling analysis to distinguish among conditions of varying sensorimotor task difficulty. The robustness was demonstrated by the consistency of such results found in the low-fidelity and a fully immersive, high-fidelity flight training environments. Moreover, the cognitive tunneling analysis revealed differences in scanning behaviour across a range of pilot expertise levels. Although more research is required to further develop this analysis to suit the needs of alternative flight scenarios, aircraft, flying conditions (i.e., VFR vs. IFR), and other applications associated with the assessment and training of gaze behaviour, the work conducted here demonstrates a promising and robust approach to provide an objective and sensitive measure of poor pilot monitoring, or as a general marker of divergence from optimal scanning.

9.2 LIMITATIONS AND FUTURE DIRECTIONS

The results presented here provide novel insights and build on previous studies examining gaze behaviour in laboratory settings and other occupational environments. However,

it is not without limitations. This section discusses several theoretical and technical limitations that should be further investigated in future research.

Five main limitations exist in the theoretical/conceptual aspects of the current findings. First, eye tracking is a method that indirectly examines changes in information processing via overt shifts in attention that are captured by gaze behaviour metrics. Our current information processing and skill performance framework is therefore limited in the extent to which it can identify and examine the covert processes that may not be associated with explicit shifts in gaze. Examples of this limitation include scenarios where covert attention is employed through peripheral vision or other sensory modalities (i.e., auditory, vestibular, kinesthetic, and tactile), which contribute additional sensory information to be processed alongside foveated visual information and can potentially impact task performance. One example of this limitation in the current work is a known ability for more experienced pilots to have expanded fields of vision that allow them to effectively process and incorporate task relevant information from their peripheral vision into subsequent actions and decision making without changing their normal gaze behaviour (Ryu et al., 2013). In a similar manner, we reported that when task manipulations were made by way of increasing cognitive load (Study 5), there were no associated alterations in gaze behaviour. Although one explanation could be that gaze behaviour is not sensitive enough to examine changes in information processing related to cognitive load, I argue that given our findings in studies two, three, and four, and the TOL task (Study 1), this explanation does not seem to be the case. I believe the experimental manipulation was too easy for the recruited pilot group, and they were not effectively challenged. This interpretation is appealing as it provides a specific use-case where such expected gaze behaviour and task performance outcomes may signal the individual has mastered a skill or reached a high level of processing and performance

efficiency. Such a result would be expected to be associated with little to no changes in task performance or gaze behaviour when alterations to the task requirements (specific to sensorimotor demands or cognitive load) are made. In any case, there are still questions about how other aspects of information processing that are not discernable through gaze behaviour may be impacted under similar conditions where gaze behaviour and performance remain unchanged despite increases in task demands. Eye tracking is only one method through which information processing can be evaluated. Thus, the examination of the interaction between information processing, task demands, and expertise should be further investigated with alternative methods that may reveal additional insight into information processing that go beyond changes in gaze behaviour including EEG, fNIRS, or even the addition of pupillometry. The latter was not examined in the current work due to limitations in eye-tracking capabilities at the time of collection. Such methods can provide more insight into the other aspects of information processing that also play a role in task performance including sensory function, motor function, covert attention, perception, and sensorimotor integration.

Second, the current work may be limited in the generalizability to a wider population. The studies presented here only involved healthy, young adults. Results, therefore, cannot be generalized to older pilots or other working populations. The lack of generalization to these older populations is a salient issue because previous work has demonstrated that older adults exhibit characteristic changes in brain structure and function associated with impairments to memory and executive function (Mirsky et al., 2011). These changes are also seen alongside age-related declines in various aspects of vision (i.e., visual acuity, scotopic sensitivity, motion sensitivity, etc.) (Anderson, 2012). As such, changes to all these critical components of information processing may modulate the effectiveness of using gaze behaviour to assess information

processing in ecological environments in the aging population. Nevertheless, the work presented here demonstrates a number of novel contributions that should be reexamined in elderly populations to determine the extent to which the current findings continue on despite the noted age-related declines in sensory and cognitive functions. Third, the current work is also limited in the generalizability to other environments, even within the context of aviation. Specifically, the current work was designed to examine the utility of gaze behaviour as a proxy for information processing in low-time pilots who are training with a single engine aircraft with a steam gauge configuration (i.e., six pack). The extent to which the reported results generalize to glass cockpits, other aircraft (i.e., more complex cockpits), IFR conditions, or real aircrafts in general is still to be determined.

Lastly, the lack of a single measure of performance in the aviation studies constrained our ability to validate the comprehensive set of gaze metrics examined here against flight performance. Unlike study one which had an overall measure of performance, the calculation of performance in the subsequent studies was based on several measures that assessed different aspects of flight performance. The absence of a singular measure of performance prevented the computation of correlations between flight performance, SA and gaze metrics across all studies. The one exception to this limitation was study two where the game provided performance scores based on landing specific performance outcomes, but it did not take into account other aircraft control parameters leading up to the aircraft touching down on the runway. Therefore, future work that aims to establish correlations with pilot flight performance should include instructor pilots in the methodology to help score flight performance and provide a singular overarching measure of proficiency for a particular task.

One main technical limitation also existed in the current work. For instance, the aspect of synchronicity across several different collection devices (i.e., task presentation device [PC, simulator], auditory tone presentation device, and eye tracker) was a constant issue with respect to the alignment of data streams for behavioural performance results and eye tracking data. Although the initiation and end of these trials were practiced and coordinated between two researchers to the best of their abilities, there is always the chance of human error and normal motor control variability in button presses that should be considered. The lack of synchronicity was particularly impactful in study five, where auditory response times could have been influenced by slight asynchronies in data collection streams. To account for some of the sources of asynchrony (i.e., lag), the calculation of response times was adjusted to take into consideration the lag between simulator start time and auditory tone initiation. Nevertheless, I believe that the evidence presented through increased error rates in auditory responses could reasonably support the general response deficits seen during increased cognitive load conditions. Notably, such a limitation also impacted the extent to which we could examine temporal contingencies between specific eye movements and actions. Therefore, future work should identify methods to aid in the synchronization of various data collection hardware to help ensure synchronicity across data streams.

Future work should build on these limitations specifically to address a growing field of questions related to aging populations, skill acquisition, and its application to extended reality environments (i.e., virtual reality, mixed reality, augmented reality, etc.). As mentioned earlier, due to the cognitive, sensory, and motor function changes that naturally occur with aging, it is important to examine the utility of using gaze metrics in this population to understand information processing. The answers generated from such research have implications for the

extent to which gaze behaviour can be used for occupational applications (i.e., re-training and assessment of the aging workforce) and for medical rehabilitation applications (i.e., treatment methods, treatment tracking and assessment, etc.). Second, the application of this research to track and assess skill acquisition is fruitful and critical in today's society where on-boarding protocols are expected to be quick and effective. What remains a crucial next step with respect to this type of application is longitudinal research that objectively tracks how gaze behaviour changes with motor skill acquisition for a given task. A lot of research has focused on the development of skill acquisition in sports, reading, and other simple motor tasks, but the role gaze behaviour plays in the acquisition of complex tasks is not yet fully established, especially in ecological environments. Therefore, the examination of how gaze behaviour characteristics change as skill development progresses will provide unique insight into task performance at different levels of skill proficiency. This insight will ultimately serve to provide more tailored feedback and improve training outcomes that have real-world implications for occupational and rehabilitation applications. Lastly, the development of cost-effective, high quality extended reality headsets with eye tracking built into the simulations of virtual, augmented, or mixed environments is proving to be a novel context for research. Therefore, future work is required to see how accurate the eye tracking and event detection algorithms are in mapping gaze behaviour in virtual space, as well as investigate how information processing in general changes within these simulated environments.

9.3 CONCLUSION

In conclusion, the work presented here provides a comprehensive assessment of the utility of using gaze behaviour measures as proxies for information processing during the

performance of a neurocognitive lab-based task and a flight simulation task within a more ecological environment. A number of gaze behaviour metrics were effective in distinguishing differences in task requirements (i.e., sensorimotor demands, cognitive load) and individual experience, which were shown to be consistent across low-fidelity and high-fidelity environments. For instance, cognitive tunneling analysis demonstrated increasing instances of longer total durations of poor pilot monitoring behaviours, particularly during more difficult task scenarios. Moreover, lower entropy values were associated with an increase in selective gaze behaviour, which was increasingly demonstrated in difficult task scenarios. Additionally, dwell time and cognitive tunneling behaviour were critical to identifying information processing mechanisms that supported successful task performance which were found to be absent in unsuccessful attempts. Taken together, the work outlined in this document provides evidence supporting the use of gaze behaviour as an effective method to objectively quantify and assess the perceptual and cognitive processes supporting skill performance. Ultimately, this work will contribute to developing training protocols that could be used to identify, correct and optimize gaze behaviour to support the learning and performance of complex tasks.

REFERENCES

- Abel, S. M. (2009). Stimulus complexity and dual tasking effects on sustained auditory attention in noise. *Aviation, space, and environmental medicine*, 80(5), 448-453.
- Abernethy, B. (1996). Training the visual-perceptual skills of athletes: Insights from the study of motor expertise. *The American Journal of Sports Medicine*, 24, S89-S92.
- Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behaviour in aviation. *Journal of Applied Research in Memory and Cognition*, 3(2), 63-71.
- Allsop, J., Gray, R., Bühlhoff, H. H., & Chuang, L. (2017). Eye movement planning on Single-Sensor-Single-Indicator displays is vulnerable to user anxiety and cognitive load. *Journal of Eye Movement Research*, 10(5), 8-1.
- Andersen, G. J. (2012). Aging and vision: changes in function and performance from optics to perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3), 403-410.
- Andrzejewska, M., & Stolińska, A. (2016). Comparing the Difficulty of Tasks Using Eye Tracking Combined with Subjective and Behavioural Criteria. *Journal of Eye Movement Research*, 9(3). <https://doi.org/10.16910/jemr.9.3.3>
- Atik, O., & Arslan, O. (2019). Use of eye tracking for assessment of electronic navigation competency in maritime training. *Journal of Eye Movement Research*, 12(3). <https://doi.org/10.16910/jemr.12.3.2>

- Ayala, N., & Heath, M. (2020). Executive Dysfunction after a Sport-Related Concussion Is Independent of Task-Based Symptom Burden. *Journal of Neurotrauma*, 37 (23), 2558-2568.
- Ayala, N., Mardanbegi, D., Zafar, A., Niechwiej-Szwedo, E., Cao, S., Kearns, S., ... & Duchowski, A. T. (2024). Does fiducial marker visibility impact task performance and information processing in novice and low-time pilots?. *Computers & Graphics*, 103889.
- Ayala, N., & Niechwiej-Szwedo, E. (2021). Effects of blocked vs. interleaved administration mode on saccade preparatory set revealed using pupillometry. *Experimental Brain Research*, 239 (1), 245-255.
- Ayala, N., Mardanbegi, D., Duchowski, A., Niechwiej-Szwedo, E., Cao, S., Kearns, S., & Irving, E. (2023a). On The Visibility Of Fiducial Markers For Mobile Eye Tracking. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* (pp. 1-7).
- Ayala, N., Zafar, A., Kearns, S., Irving, E., Cao, S., & Niechwiej-Szwedo, E. (2023b). The effects of task difficulty on gaze behaviour during landing with visual flight rules in low-time pilots. *Journal of Eye Movement Research*.
- Ayala, N., Zafar, A., & Niechwiej-Szwedo, E. (2022). Gaze behaviour: a window into distinct cognitive processes revealed by the Tower of London test. *Vision Research*.
- Babu, M. D., JeevithaShree, D. V., Prabhakar, G., Saluja, K. P. 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).

- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829-839.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioural and Brain Sciences*, 20(4), 723-742.
- Batty, M., Morphet, R., Masucci, P., & Stanilov, K. (2014). Entropy, complexity, and spatial information. *Journal of Geographical Systems*, 16(4), 363-385.
- Beall, A. C., & Loomis, J. M. (1997). Optic flow and visual analysis of the base-to-final turn. *The International Journal of Aviation Psychology*, 7(3), 201-223.
- Bell, M. A., Facci, E. L., & Nayeem, R. V. (2005). Cognitive Tunneling, Aircraft-Pilot Coupling Design Issues and Scenario Interpretation Under Stress in Recent Airline Accidents. 2005 *International Symposium on Aviation Psychology*, 45-49.
- Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, 68, 569-579.
- Bhojwani, T. M., Lynch, S. D., Bühler, M. A., & Lamontagne, A. (2022). Impact of dual tasking on gaze behaviour and locomotor strategies adopted while circumventing virtual pedestrians during a collision avoidance task. *Experimental brain research*, 240(10), 2633-2645.
- Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, 33, 1-21.

- Boeing. (2021). Statistical Summary of Commercial Jet Airplane Accidents-Worldwide Operations 1959-2021. https://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf.
- Brams, S., Hooge, I. T., Ziv, G., Dauwe, S., Evens, K., De Wolf, T., ... & Helsen, W. F. (2018). Does effective gaze behaviour lead to enhanced performance in a complex error-detection cockpit task?. *PloS one*, 13(11), e0207439.
- Brams, S., Ziv, G., Levin, O., Wagemans, J., Williams, A. M., & Helsen, W. F. (2020). Brain, gaze behaviour and perceptual-cognitive skills in aviation: what is yet to be studied?. In *Eye-Tracking in Aviation. Proceedings of the 1st International Workshop (ETAVI 2020)* (pp. 100-107). ISAE-SUPAERO, Université de Toulouse; Institute of Cartography and Geoinformation (IKG), ETH Zurich.
- Brown, D. L., Vitense, H. S., Wetzel, P. A., & Anderson, G. M. (2002). Instrument scan strategies of F-117A Pilots. *Aviation, Space, and Environmental Medicine*, 73(10), 1007-1013.
- Burris, K., Liu, S., & Appelbaum, L. (2020). Visual-motor expertise in athletes: Insights from semiparametric modelling of 2317 athletes tested on the Nike SPARQ Sensory Station. *Journal of Sports Sciences*, 38(3), 320-329.
- Cao, S., Samuel, S., Murzello, Y., Ding, W., Zhang, X., & Niu J. (2022). Hazard perception in driving: A systematic literature review. *Transportation Research Record*. <https://doi.org/10.1177/03611981221096666>

- Cao, S., & Liu, Y. (2013). Concurrent processing of vehicle lane keeping and speech comprehension tasks. *Accident Analysis & Prevention*, 59, 46-54.
- Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3), 6-6.
- Ciuperca, G., & Girardin, V. (2007). Estimation of the entropy rate of a countable Markov chain. *Communications in Statistics—Theory and Methods*, 36(14), 2543-2557.
- Clark, J. J. (1999). Spatial attention and latencies of saccadic eye movements. *Vision Research*, 39(3), 585-602.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201-215.
- Crosby, J. V., & Parkinson, S. R. (1979). A dual task investigation of pilots' skill level. *Ergonomics*, 22(12), 1301–1313. <https://doi.org/10.1080/00140137908924704>
- Cvahte Ojsteršek, T., & Topolšek, D. (2019). Eye tracking use in researching driver distraction: A scientometric and qualitative literature review approach. *Journal of Eye Movement Research*, 12(3). <https://doi.org/10.16910/jemr.12.3.5>
- de Brouwer, A. J., Flanagan, J. R., & Spering, M. (2021). Functional use of eye movements for an acting system. *Trends in Cognitive Sciences*.
- Dehais, F., Behrend, J., Peysakhovich, V., Causse, M., & Wickens, C. D. (2017). Pilot flying and pilot monitoring's aircraft state awareness during go-around execution in aviation: A behavioural and eye tracking study. *The International Journal of Aerospace Psychology*, 27(1-2), 15-28.

- Dehais, F., Juaneda, S., & Peysakhovich, V. (2020). Monitoring eye movements in real flight conditions for flight training purpose. *ETAVI '20- Proceedings of the 1st International Workshop on Eye-Tracking in Aviation*, 6, 54-60.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-168.
- Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L. J., Catena, A., & Di Stasi, L. L. (2019). The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. *Applied Ergonomics*, 77, 92-99.
- Dick, A. O. (1976). Spatial abilities. In *Studies in Neurolinguistics* (225-268). Academic Press.
- Dick, A. O. (1980). Instrument scanning and controlling: using eye movement data to understand pilot behaviour and strategies. *Washington, DC: National Aeronautics and Space Administration, NASA CR-3306*.
- 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.
- Endsley, M.R. (1988) Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting*, 32, 97-101.
- Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human factors*, 37(1), 65-84.
- Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation research part F: traffic psychology and behaviour*, 8(2), 97-120.

- Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human factors*, 59(5), 734-764.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, 7(2), 336.
- Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in cognitive sciences*, 10(8), 382-390.
- Federal Aviation Administration (2021). Airplane Flying Handbook. *Washington, DC: Federal Aviation Administration* (FAA-H-8083-3C).
- Flanagan, J. R., Terao, Y., & Johansson, R. S. (2008). Gaze behaviour when reaching to remembered targets. *Journal of Neurophysiology*, 100(3), 1533-1543.
- Garrison, T. M., & Williams, C. C. (2013). Impact of relevance and distraction on driving performance and visual attention in a simulated driving environment. *Applied cognitive psychology*, 27(3), 396-405.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523-552.
- Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. *Journal of Eye Movement Research*, 6(1).
<https://doi.org/10.16910/jemr.6.1.3>

- Glaholt, M. G. (2014). Eye tracking in the cockpit: a review of the relationships between eye movements and the aviators cognitive state. <https://apps.dtic.mil/sti/pdfs/AD1000097.pdf>.
- Gray, R., Navia, J. A., & Allsop, J. (2014). Action-specific effects in aviation: What determines judged runway size?. *Perception*, 43(2-3), 145-154.
- Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, 33(5), 623-642.
- Gonzalez, D. A., & Niechwiej-Szwedo, E. (2016). The effects of monocular viewing on hand-eye coordination during sequential grasping and placing movements. *Vision Research*, 128, 30-38.
- Hancock, P. A. (2009). The future of simulation. In *Human factors in simulation and training* (pp. 297-320). CRC Press.
- Harris, R. L., Glover, B. L., & Spady, A. A. (1986). Analytical techniques of pilot scanning behaviour and their application. *Washington, DC: National Aeronautics and Space Administration, NASA TP- 2525*.
- Hebbar, A., Vinod, S., A.K., S., Pashilkar, A., & Biswas, P. (2023). Cognitive load estimation in VR flight simulator. *Journal of Eye Movement Research*, 15(3).
<https://doi.org/10.16910/jemr.15.3.11>
- Hermens, F., Flin, R., & Ahmed, I. (2013). Eye movements in surgery : A literature review. *Journal of Eye Movement Research*, 6(4). <https://doi.org/10.16910/jemr.6.4.4>

- Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower-of-London task. *Journal of Cognitive Neuroscience*, 12(5), 894–907. <https://doi.org/10.1162/089892900562499>
- Hodgson, T. L., Ezard, G., & Hermens, F. (2019). Eye Movements in Neuropsychological Tasks. *Current Topics in Behavioural Neurosciences*, 41, 393–418. https://doi.org/10.1007/7854_2019_98.
- Hodgson, T. L., Tiesman, B., Owen, A. M., & Kennard, C. (2002). Abnormal gaze strategies during problem solving in Parkinson's disease. *Neuropsychologia*, 40(4), 411-422.
- Huddy, V. C., Hodgson, T. L., Kapasi, M., Mutsatsa, S. H., Harrison, I., Barnes, T. R., & Joyce, E. M. (2007). Gaze strategies during planning in first-episode psychosis. *Journal of Abnormal Psychology*, 116(3), 589.
- Hudspeth, A. J., Jessell, T. M., Kandel, E. R., Schwartz, J. H., & Siegelbaum, S. A. (Eds.). (2013). *Principles of neural science*. McGraw-Hill, Health Professions Division.
- Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye* (pp. 573-605). North-Holland.
- Kaller, C. P., Rahm, B., Bolkenius, K., & Unterrainer, J. M. (2009). Eye movements and visuospatial problem solving: Identifying separable phases of complex cognition. *Psychophysiology*, 46(4), 818-830.
- Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment*, 24(1), 46.

- Kandel, E., Schwartz, J., Jessell, T., Siegelbaum, S., & Hudspeth, A. J. (2012) Principles of Neural Science (5th ed). McGraw-Hill Medical.
- Kanaan, D., & Moacdieh, N. M. (2022). Eye Tracking to Evaluate the Effects of Interruptions and Workload in a Complex Task. *Human Factors*, 64(7), 1168-1180.
- Kearns, S.K. (2021). Fundamentals of International Aviation (2nd ed.). Routledge.
- Kim, J., Palmisano, S. A., Ash, A., & Allison, R. S. (2008). Pilot gaze and glideslope control. *ACM Transactions on Applied Perception (TAP)*, 7(3), 1-18.
- Krebs, M. J., & Wingert, J. W. (1976). Use of the oculometer in pilot workload measurement. *Washington, DC: National Aeronautics and Space Administration, NASA CR-144951*.
- Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 159-166).
- Kübler, T. C., Kasneci, E., Rosenstiel, W., Aehling, K., Heister, M., Nagel, K., Schiefer, U., & Papageorgiou, E. (2015). Driving with Homonymous Visual Field Defects: Driving Performance and Compensatory Gaze Movements. *Journal of Eye Movement Research*, 8(5). <https://doi.org/10.16910/jemr.8.5.5>
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, 9(2), 75-82.
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities?. *Vision Research*, 41(25-26), 3559-3565.

- León, J. A., Moreno, J. D., Escudero, I., & Kaakinen, J. K. (2019). Selective attention to question-relevant text information precedes high-quality summaries: Evidence from eye movements. *Journal of Eye Movement Research*, 12(1).
<https://doi.org/10.16910/jemr.12.1.6>
- Lemonnier, S., Brémond, R., & Baccino, T. (2014). Discriminating cognitive processes with eye movements in a decision-making driving task. *Journal of Eye Movement Research*, 7(4).
<https://doi.org/10.16910/jemr.7.4.3>
- Lijing, W., & Lin, C. (2016). Eye movement comparison of professional and novice pilots in key actions of takeoff phase. *2016 IEEE International Conference on Aircraft Utility Systems (AUS)*, pp703-706.
- International Air Transport Association (2023). *Competency Assessment and Evaluation for Pilots, Instructors, and Evaluators*. GUIDANCE MATERIAL.
- Liu, S., Donaldson, R., Subramaniam, A., Palmer, H., Champion, C. D., Cox, M. L., & Appelbaum, L. G. (2021). Developing expert gaze pattern in laparoscopic surgery requires more than behavioural training. *Journal of Eye Movement Research*, 14(2).
<https://doi.org/10.16910/jemr.14.2.2>
- Lounis, C., Peysakhovich, V., & Causse, M. (2021). Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study. *PLoS One*, 16(2), e0247061.
- Lu, T., Lou, Z., Shao, F., Li, Y., & You, X. (2020). Attention and Entropy in Simulated Flight with Varying Cognitive Loads. *Aerospace Medicine and Human Performance*, 91(6), 489-495.

- Lutnyk, L., Rudi, D., Schinazi, V. R., Kiefer, P., & Raubal, M. (2023). The effect of flight phase on electrodermal activity and gaze behaviour: A simulator study. *Applied Ergonomics*, 109, 103989.
- Manor, B., Costa, M. D., Hu, K., Newton, E., Starobinets, O., Kang, H. G., ... & Lipsitz, L. A. (2010). Physiological complexity and system adaptability: evidence from postural control dynamics of older adults. *Journal of Applied Physiology*, 109(6), 1786-1791.
- Martin, L., Tapper, A., Gonzalez, D. A., Leclerc, M., & Niechwiej-Szwedo, E. (2017). The effects of task-relevant saccadic eye movements performed during the encoding of a serial sequence on visuospatial memory performance. *Experimental Brain Research*, 235(5), 1519.
- Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14(2), 83-96.
- Menekşe Dalveren, G. G., & Cagiltay, N. E. (2018). Using Eye-Movement Events to Determine the Mental Workload of Surgical Residents. *Journal of Eye Movement Research*, 11(4). <https://doi.org/10.16910/jemr.11.4.3>
- Mertens, H. W., & Lewis, M. F. (1981). *Effect of different runway size on pilot performance during simulated night landing approaches*. FEDERAL AVIATION ADMINISTRATION WASHINGTON DC OFFICE OF AVIATION MEDICINE.
- Mirsky, J. B., Heuer, H. W., Jafari, A., Kramer, J. H., Schenk, A. K., Viskontas, I. V., ... & Boxer, A. L. (2011). Anti-saccade performance predicts executive function and brain structure in normal elders. *Cognitive and Behavioural Neurology*, 24(2), 50-58.

- National Transportation Safety Board (1994). A Review of Flightcrew-Involved, Major Accidents of US Air Carriers, 1978 Through 1990. https://reports.aviation-safety.net/1994/19940618-1_LJ25_XA-BBA.pdf.
- Nibbeling, N., Oudejans, R. R., & Daanen, H. A. (2012). Effects of anxiety, a cognitive secondary task, and expertise on gaze behaviour and performance in a far aiming task. *Psychology of Sport and Exercise*, 13(4), 427-435.
- Nitschke, K., Ruh, N., Kappler, S., Stahl, C., & Kaller, C. P. (2012). Dissociable stages of problem solving (I): Temporal characteristics revealed by eye-movement analyses. *Brain and Cognition*, 80(1), 160-169.
- Niu, Y.- feng, Gao, Y., Xue, C.- qi, Zhang, Y.- ting, & Yang, L.- xin. (2020). Improving eye–computer interaction interface design: Ergonomic investigations of the optimum target size and gaze-triggering dwell time. *Journal of Eye Movement Research*, 12(3).
<https://doi.org/10.16910/jemr.12.3.8>
- Noble, C. (2002). The Relationship Between Fidelity And Learning In Aviation Training And Assessment. *Journal of Air Transportation*, 7(3).
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behaviour research methods*, 42(1), 188-204.
- Paas, F., & van Merriënboer, J. J. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, 29(4), 394-398.

- Panchuk, D., & Vickers, J. N. (2006). Gaze behaviours of goaltenders under spatial–temporal constraints. *Human Movement Science*, 25(6), 733-752.
- 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.
- Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139(3), 266-277.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.
- Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 13-19.
- Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5-21.
- Robinski, M., & Stein, M. (2013). Tracking Visual Scanning Techniques in Training Simulation for Helicopter Landing. *Journal of Eye Movement Research*, 6(2).
<https://doi.org/10.16910/jemr.6.2.3>
- Rowe, J. B., Owen, A. M., Johnsrude, I. S., & Passingham, R. E. (2001). Imaging the mental components of a planning task. *Neuropsychologia*, 39(3), 315-327.
- Ryu, D., Abernethy, B., Mann, D. L., Poolton, J. M., & Gorman, A. D. (2013). The role of central and peripheral vision in expert decision making. *Perception*, 42(6), 591-607.

- Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioural and eye-tracking data. *Human Factors*, 49(3), 347-357.
- Schwenk, M., Zieschang, T., Oster, P., & Hauer, K. (2010). Dual-task performances can be improved in patients with dementia: a randomized controlled trial. *Neurology*, 74(24), 1961-1968.
- Selcon, S. J., Taylor, R. M., & Koritsas, E. (1991, September). Workload or situational awareness?: TLX vs. SART for aerospace systems design evaluation. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 35, No. 2, pp. 62-66). Sage CA: Los Angeles, CA: SAGE Publications.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 298(1089), 199-209.
- Shannon, C. E. (1984). A mathematical theory of communication. *The Bell System Technical Journal*, 27 (3), 379-423.
- Shappell, S. A., & Wiegmann, D. A. (2000). The human factors analysis and classification system--HFACS.
- Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioural Reviews*, 96, 353-366.
- Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M., Berlowitz, D. J., ... & Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Scientific Reports*, 8(1), 1-10.

- Spady Jr, A. A. (1978). Airline pilot scan patterns during simulated ILS approaches. *Washington, DC: National Aeronautics and Space Administration, NASA TP-1250.*
- Stergiou, N., & Decker, L. M. (2011). Human movement variability, nonlinear dynamics, and pathology: is there a connection?. *Human movement science*, 30(5), 869-888.
- Stuhr, C., Hughes, C. M. L., & Stöckel, T. (2018). Task-specific and variability-driven activation of cognitive control processes during motor performance. *Scientific reports*, 8(1), 10811.
- Sun, Q., Xia, J. (Cecilia), Falkmer, T., & Lee, H. (2016). Investigating the Spatial Pattern of Older Drivers' Eye Fixation Behaviour and Associations with Their Visual Capacity. *Journal of Eye Movement Research*, 9(6). <https://doi.org/10.16910/jemr.9.6.2>
- Svensson, E., Angelborg-Thanderez, M., Sjöberg, L., & Olsson, S. (1997). Information complexity-mental workload and performance in combat aircraft. *Ergonomics*, 40(3), 362-380.
- Tole, J. R., Stephens, A. T., Harris, R. L., & Ephrath, A. R. (1982). Visual scanning behaviour and mental workload in aircraft pilots. *Aviation, Space, and Environmental Medicine*. 53(1), 54-61.
- Tang, H., Day, E., Kendhammer, L., Moore, J. N., Brown, S. A., & Pienta, N. J. (2016). Eye Movement Patterns in Solving Science Ordering Problems. *Journal of Eye Movement Research*, 9(3). <https://doi.org/10.16910/jemr.9.3.6>
- Taylor, R. M., & Selcon, S. J. (1990). Cognitive Quality and Situational Awareness with Advanced Aircraft Attitude Displays. *Proceedings of the Human Factors Society Annual Meeting*, 34(1), 26-30.

- Tole, J. R., Stephens, A. T., & Vivaudou, E. A. & Young, LR (1983). Visual scanning behaviour and pilot workload. *Washington, DC: National Aeronautics and Space Administration, NASA CR-3717.*
- Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movements. *Psychonomic Bulletin & Review, 13*(3), 452-457.
- van De Merwe, K., Van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology, 22*(1), 78-95.
- van Dijk, H., van de Merwe, K., & Zon, R. (2011). A Coherent Impression of the Pilots' Situation Awareness: Studying relevant human factors tools. *The International Journal of Aviation Psychology, 21*(4), 343-356.
- van Leeuwen, P. M., Happee, R., & de Winter, J. C. (2015). Changes of driving performance and gaze behaviour of novice drivers during a 30-min simulator-based training. *Procedia Manufacturing, 3*, 3325-3332.
- Vansteenkiste, P., Lenoir, M., Krejtz, I., & Krejtz, K. (2022). Visual strategies of young soccer players during a passing test – A pilot study. *Journal of Eye Movement Research, 15*(1). <https://doi.org/10.16910/jemr.15.1.3>
- Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In *Proceedings of the XXVII Conference of the Cognitive Science Society*, 1.

- Veneri, G., Rosini, F., Federighi, P., Federico, A., & Rufa, A. (2012). Evaluating gaze control on a multi-target sequencing task: the distribution of fixations is evidence of exploration optimisation. *Computers in Biology and Medicine*, 42(2), 235-244.
- Vickers, J. N. (2007). *Perception, cognition, and decision training: The quiet eye in action*. Human Kinetics.
- Vickers, J. N., & Williams, A. M. (2017). The role of mental processes in elite sports performance. In *Oxford Research Encyclopedia of Psychology*.
- Vlačić, S. I., Knežević, A. Z., Mandal, S., Rođenkov, S., & Vitsas, P. (2020). Improving the pilot selection process by using eye-tracking tools. *Journal of Eye Movement Research*, 12(3). <https://doi.org/10.16910/jemr.12.3.4>
- Waller, M. C. (1976). An investigation of correlation between pilot scanning behaviour and workload using stepwise regression analysis. *Washington, DC: National Aeronautics and Space Administration, NASA TM X-3344*.
- Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour*, 26, 227-237.
- Weiss, R. S., Remington, R., & Ellis, S. R. (1989). Sampling distributions of the entropy in visual scanning. *Behaviour Research Methods, Instruments, & Computers*, 21(3), 348-352.

- 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
- Wright, J. L., Quinn, S. A., Chen, J. Y. C., & Barnes, M. J. (2014). Individual Differences in Human-Agent Teaming: An Analysis of Workload and Situation Awareness through Eye Movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 1410-1414. <https://doi.org/10.1177/1541931214581294>
- Xiong, W., Wang, Y., Zhou, Q., Liu, Z., & Zhang, X. (2016). The research of eye movement behaviour of expert and novice in flight simulation of landing. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, 485-493.
- Yang, S., Kuo, J., & Lenné, M. G. (2018). Analysis of gaze behaviour to measure cognitive distraction in real-world driving. *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 62, 1944-1948.
- Yang, S., Wilson, K. M., Roady, T., Kuo, J., & Lenné, M. G. (2022). Evaluating driver features for cognitive distraction detection and validation in manual and Level 2 automated driving. *Human factors*, 64(4), 746-759.
- Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211). Springer, Boston, MA.
- Zafar, A., Martin Calderon, C., Yeboah, A. M., Dalton, K., Irving, E., & Niechwiej-Szwedo, E. (2023). Investigation of Camera-Free Eye-Tracking Glasses Compared to a Video-Based System. *Sensors*, 23(18), 7753.

- Zhou, J., Hao, Y., Wang, Y., Jor'dan, A., Pascual-Leone, A., Zhang, J., ... & Manor, B. (2014). Transcranial direct current stimulation reduces the cost of performing a cognitive task on gait and postural control. *European Journal of Neuroscience*, 39(8), 1343-1348.
- Ziv, G. (2016). Gaze behaviour and visual attention: A review of eye tracking studies in aviation. *The International Journal of Aviation Psychology*, 26(3-4), 75-104.
- Ziv, G., & Lidor, R. (2016). The effect of age on gaze behaviour in drivers and pedestrians – a review. *Journal of Eye Movement Research*, 9(7). <https://doi.org/10.16910/jemr.9.7.6>
- Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel– serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 244.